

UNIVERSITA' DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA INDUSTRIALE

CORSO DI LAUREA IN INGEGNERIA CHIMICA E DEI PROCESSI INDUSTRIALI

**Tesi di Laurea Magistrale in
Ingegneria Chimica e dei Processi Industriali**

**OPTIMAL DESIGN OF A MINIMUM SET OF CLINICAL TESTS
FOR THE IDENTIFICATION AND CHARACTERIZATION OF
VON WILLEBRAND DISEASE**

Relatore: Prof. Fabrizio Bezzo

Correlatore: Dott. Federico Galvanin

Laureanda: BEATRICE TAVERNA

ANNO ACCADEMICO 2016-2017

*Ai miei genitori,
per tutti i sacrifici che hanno fatto per portarmi a questo,
per il loro costante sostegno,
per la serenità che mi hanno sempre regalato,
per il bene immenso che io voglio loro,
Grazie.*

Abstract

Von Willebrand disease (VWD) is one of the most severe inherited bleeding disorder recognizable in humans under different typologies, characterized by qualitative or quantitative deficiencies of von Willebrand factor (VWF). The diagnosis of the disease is difficult and linked to the experience of doctors, therefore a mechanistic model by Galvanin et al. has recently been proposed and applied to help in the disease characterization and diagnosis. However, model identification requires the execution of the stressful and invasive DDAVP test. The test can cause severe consequences on the patients and it cannot be carried out on babies or old people. Therefore, it appears clearly reasonable to study a way for identifying the model without the DDAVP test but only with basal clinical trials. This research work pretends to be only the first step towards achieving the ultimate just mentioned target. Response surface methodology (RSM) has been applied to find explicit correlations for the elimination constant k_e , the proteolytic kinetic constant k_1 and the constant of release k_0 as function of basal quantities (VWF:R, VWFpp ratio, VWF:Ag), whose values are derived from a simple blood sample. Results show that, the new equations, once substituted in the PK model of VWD by Galvanin et al., allow for the identification of the modified model and the simulated responses demonstrate that the predictive capability of the model is fully maintained.

Contents

Introduction	1
Introduction to von Willebrand disease	3
1.1 Von Willebrand disease: alteration of von Willebrand factor in the coagulation process	3
1.2 VWD categories	5
1.2.1 Type 1	5
1.2.2 Type 3	6
1.2.3 Type 2A.....	6
1.2.4 Type 2B.....	6
1.2.5 Type 2N.....	7
1.2.6 Type Vicenza	7
1.3 Diagnosis and detection of VWD.....	8
1.3.1 Clinical tests	9
1.4 Pharmacokinetic model of VWD	10
1.5 Objectives.....	11
Identification and modification of VWD pharmacokinetic model	13
2.1 The SGM mathematical structure	13
2.1.1 Parameters estimation procedure	14
2.2 Methodology	15
2.2.1 Model-based design of experiments (MBD _{oE})	15
2.2.2 Parameter estimation problem	17
2.2.3 MBD _{oE} for parameter estimation.....	23
2.2.4 Response surface methodology (RSM).....	24
2.4.6 Software	27
2.3 Model modification and identification approach	27
Preliminary analysis of the model information content	29
3.1 Results of the sensitivity analysis	29

3.2 Results of FIM (Fisher Information Matrix) evaluation	34
Model modification and validation	37
4.1 Results of model modification	37
4.1.1 RSM for k_e	38
4.1.1.1 SGM modification and simulation of MGM_1.....	45
4.1.1.2 Information content analysis on MGM_1	48
4.1.2 RSM for k_1	53
4.1.2.1 SGM (2017) modification and simulation of MGM_2	57
4.1.2.2 Information content analysis on MGM_2.....	61
4.1.3 RSM for k_0	64
4.1.3.1 SGM modification and simulation of MGM_3.....	72
4.2 Model validation.....	79
4.3 Redesign of the DDAVP clinical trial.....	84
4.4 Model representation of out-of-category subjects	95
Symbols.....	101
Acknowledgments.....	103
References	105
Appendix A	107
A.1 Available dataset for healthy subjects	107
A.2 Available dataset for 2B and Vicenza subjects	108
Appendix B	109
B.1 Investigation of the response surface for parameter k_e	109
B.2 Investigation of the response surface for parameter k_1	113
B.3 Investigation of the response surface for parameter k_0	118
Appendix C	123
C.1 In silico experimental data	123
C.2 Punctual sampling time optimization results.....	128

Introduction

The work of thesis has been carried out at University College London (UCL) in collaboration with the Hospital of Padova.

The overall objective of this thesis is to optimally design a minimal set of basal clinical trials for the identification of the pharmacokinetic (PK) model of von Willebrand disease (VWD) in order to decrease the time and effort required for the disease characterization and diagnosis.

VWD is one of the most diffuse bleeding disorders visible in humans, caused by an alteration of von Willebrand factor (VWF), a multimeric glycoprotein present in the blood stream. VWF acts a fundamental function in the haemostatic process and its alteration reflects into a coagulation disorder. VWD occurs in a large variety of forms and its symptoms range from sporadic nosebleeds and mild bleeding from small lesions in skin, to acute thrombocytopenia or prolonged bleeding episodes. Diagnosis of VWD may be complicated due to the various number of VWD types (1, 2A, 2B, Vicenza), therefore pharmacokinetic models have been recently proposed for the classification of the disease, elucidating the critical pathways involved in the disease characterization. However, the complexity of the models requires long (at least 24 h) and invasive non-routine tests like DDAVP to be carried out on the subjects to achieve a statistically satisfactory estimation of the individual metabolic parameters. Therefore, the scientific community is pushing researchers to study a way for identifying the mechanistic model of VWD without the need of DDAVP test, exploiting only basal clinical trials. The alternative basal tests which are considered in this study are: (a) Propeptide test (VWFpp) to quantify VWF elimination from the blood stream; (b) Antigen VWF test (VWF:Ag) to evaluate the number of VWF antigens in the bloodstream; (c) Collagen-binding VWF test (VWF:CB) to analyse the ability of VWF in binding with collagen. From these clinical trials two other important physiological quantities are derived for diagnostic purposes: *i*) VWFpp ratio expressed as the ratio between VWFpp and VWF:Ag; *ii*) VWF:R defined as the ratio between VWF:CB and VWF:Ag.

To achieve the overall target of the work, a simplified version of the mechanistic model of VWD by Galvanin et al. is used. The model assumes that high (HMW) and ultrahigh (ULMW) molecular weight multimers are released in the bloodstream from the endothelial cells. Then, HMW and ULMW multimers are cleaved by the metalloproteinase ADAMTS13 into low molecular weight multimers (LMW) and eliminated from the bloodstream. A modified model is proposed in this work by including in the formulation a set of explicit correlations linking the PK parameters (k_0 , k_1 , k_e) to basal VWF:Ag, VWF:CB and VWFpp clinical trials and readings. The new equations are obtained using response surface methodology (RSM). RSM is a design of experiment technique, used to develop

“black-box” regression models to establish a correlation between inputs and outputs in a system. The approach is used to approximate the information coming from experimental data with the aim of defining the profile of responses in the experimental design space.

A two-stage model identification procedure is applied based on RSM. In the first step, a model discrimination based on Akaike index is used to determine the best structure of the response surface (linear, quadratic, with interactions, etc.). In the second step a data mining procedure is carried out to estimate RSM parameters in the most precise way. Following the procedure suggested by Asprey and Macchietto, identifiability tests (sensitivity analysis and information content analysis) are conducted to evaluate if PK model parameters can be uniquely estimated from the experimental data. Linear response surfaces with interactions have been successfully applied for each type of VWD and it represents the first step for finding explicit correlations between the kinetic parameters (k_0 , k_1 , k_e) and the three basal trials: VWF:Ag, VWF:CB, VWFpp and their related quantities.

Explicit correlations have been found for k_1 and k_e with high degree of accuracy. Results show a good agreement between the simplified and the modified model including RSM correlations.

A non-negligible degree of uncertainty is instead obtained when also an explicit correlation for k_0 is added to the equation set. This finding is confirmed also by the medical community which stated that the VWF release path is the most critical to define with accuracy.

The modified model with the explicit correlation for k_1 and k_e has been then used to redesign the DDAVP clinical trial. As result the DDAVP assay has been reduced to 3 hours instead of 24 hours. This strong reduction is possible because only the release kinetic constant k_0 needs now to be estimated. Parameter k_0 is linked to the release from the endothelial cells, which represents the first step of the VWF path in the bloodstream.

The work of thesis represents a first step towards reaching model identification starting only from basal clinical trials. This will allow an accurate characterization of the disease without the invasive DDAVP test, resulting therefore into a better quality of life for the patients.

Chapter 1

Introduction to von Willebrand disease

In this chapter, a general overview on the disease and clinical management is given. VWD categories and the diagnostic instruments are described to stress out the difficulties that the operators meet in characterising the disease and how a model-based approach can help them in the diagnostic process.

1.1 Von Willebrand disease: alteration of von Willebrand factor in the coagulation process

Von Willebrand disease (VWD) is one of the most common bleeding disorders visible in humans, discovered by the Finnish physician Eric von Willebrand in 1926 (Berntrop, 2007). Von Willebrand disease is caused by a qualitative or quantitative deficiency of von Willebrand factor (VWF), that is a multimeric glycoprotein composed by a variable number of identical subunits, consisting of 2050 aminoacid residues and up to 22 carbohydrate side chains (Zaverio, 2007). VWF multimers vary between dimers of 225,000 dalton and large structures consisting of more than 50 subunits. VWF is produced by endothelial cells and megakaryocytes and it can be found in subendothelial matrix, blood plasma and platelets. Considering its synthesis, the VWF produced in the endothelial cells can be secreted through two different pathways: a constitutive pathway, in which molecules are immediately released after synthesis and a regulated pathway where storage in organelles called Weibel-Palade (in endothelial cells) and α -granules (in megakaryocytes) is involved. The presence of uncleaved large multimers is fundamental to guarantee action where platelet adhesion and aggregation is necessary. The initially uncleaved VWF is then subjected to reduction in size through a proteolytic cleavage controlled by the metalloproteinase ADAMTS-13. Various size of the circulating multimers means different proadhesive functions. In particular the largest VWF multimers show enhanced thrombogenic functions, indeed they are released at the time of injury where the tissue is damaged, but in order to prevent excessive thrombus formation the physiological regulatory mechanism causes their cleavage from the circulation (Zaverio, 2007). VWF plays a fundamental role in the coagulation process: in fact, it mediates platelet adhesion, platelet aggregation, thrombus growth, and it binds, transports and protects the coagulation factor VIII. Precisely, in the haemostatic process, VWF attaches to subendothelial collagen at the one side and to platelets at the glycoprotein Ib receptor at

the other side, therefore a platelet plug can be formed in the presence of vessel injuries. The VWF protein is encoded by a gene on human chromosome 12 and its primary structure, reported in figure 1.1, shows several repeated domains.

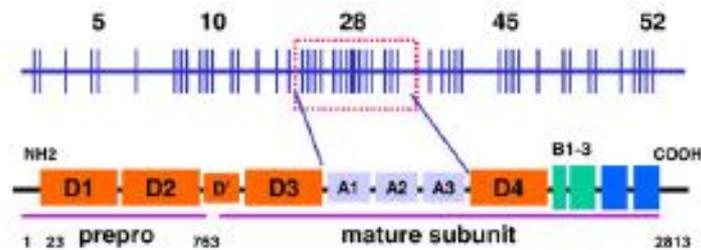


Figure 1.1. Primary structure of VWF protein (Lillicrap, 2007).

The D1, D2, D' and D3 domains are involved in the regulation of the process of multimers formation, D' and D3 are also linked in the FVIII binding, while A1 and A3 domains possess collagen-binding properties. Precisely A1 contains the exclusive binding site for the platelet with receptor glycoprotein (GP) α 1b, whereas A3 is the domain through which VWF binds to collagen (Lillicrap, 2007).

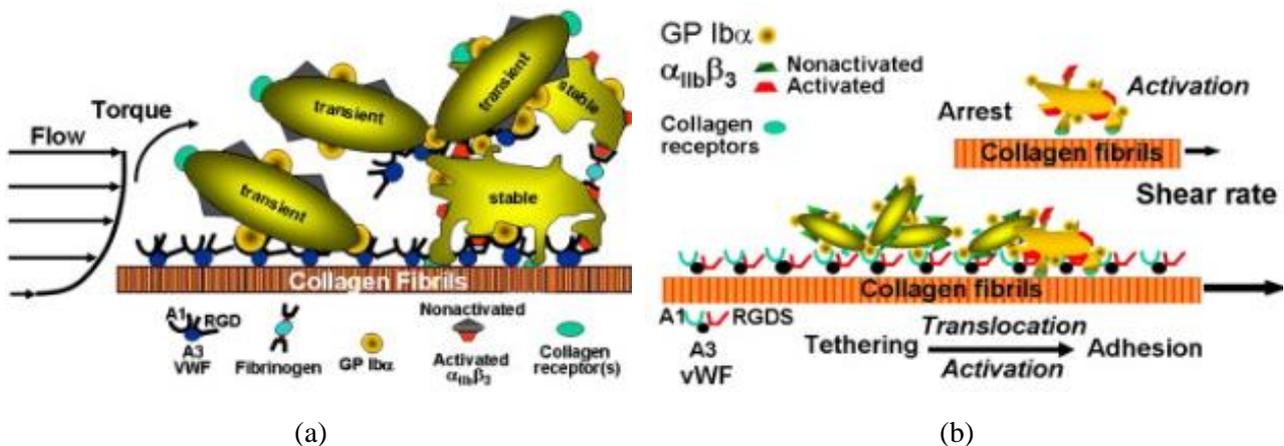


Figure 1.2. Platelets adhesion (a) and aggregation (b) (Lillicrap, 2007).

The structure of von Willebrand factor allows its binding to collagen and the interaction with specific receptors on the membrane of platelets to ensure the initiation and propagation of the coagulation process. As can be seen from figure 1.2.a, subendothelial VWF is able to bind platelets through glycoprotein α 1b in case of vascular injury also in vessel characterized by high shear stresses. Hence, the movement of platelets is significantly reduced until platelets are able to adhere stably. Then, activated platelets secrete the contents of their granules and bind adhesive proteins from plasma, such

as fibrinogen and VWF, which form the substrate where platelets aggregation (figure 1.2.b) can develop to form thrombus.

Von Willebrand disease can clearly be defined as an alteration of VWF in the coagulation process, indeed the predominant clinical symptoms are nosebleeds, bleeding from small lesions in skin, mucosa or the gastrointestinal tract, menorrhagia, excessive bleeding after trauma, surgical intervention or childbirth. There are several types of VWD (1, 2A, 2B, 2M, 2N, Vicenza, 3), but the central feature of all forms of VWD is the presence of reduced amounts of VWF or abnormal forms of VWF in the circulation. Precisely, Type 1 and 3 are characterized by mild or severe quantitative deficiencies of VWF, respectively, on the other hand, type 2 variants are characterised by qualitative deficiencies caused by mutations in the VWF gene (Sadler, 2003).

1.2 VWD categories

Considering some epidemiologically data, the presence of von Willebrand disease is about one in 100 individuals, but most of them are asymptomatic. The prevalence of clinically significant cases is therefore one in 10000 individuals. Particularly, it is mainly detected in women and apparently more severe forms appear in people with blood type 0 (Lillicrap, 2007).

Von Willebrand disease appears in several forms. In order to understand the diagnostic issues, a brief description of the most important VWD types is given below.

1.2.1 Type 1

Type 1 VWD is the most common form with almost 80% of the total cases. It is characterized by a mild to moderate quantitative deficiency of VWF but its functionality is normal. The inheritance is dominant, but the penetrance is strongly variable, indeed given the specific genotype one person can be asymptomatic having normal clinical tests, while another can have mild or moderate symptoms with some abnormal clinical tests. Therefore, the detection of this typology of disease is really challenging. The diagnostic problem is linked to the presence of an arbitrary threshold, that separates normal from abnormal VWF levels, indeed the boundary between type 1 VWD and low VWF should be better defined by determining the likelihood of an intragenic VWF mutation as a function of VWF level. Another issue, which complicates the diagnosis, is that the normal range of VWF levels is really broad, in fact 95% of the values fall between 50% and 200% of the means. Again, VWF level is influenced by ABO blood type with the 0 group characterized by the lowest quantity. Moreover, the symptoms which characterize VWD type 1 are mainly common and medically insignificant, so they represent a necessary but not sufficient condition for the diagnosis. All the problems just described

lead unfortunately to misdiagnosis and, indeed, many people diagnosed with VWD type 1 do not have the disease at all. Therefore, extremely negative consequences arise. Many false-positive patients are subjected to risky, expensive and useless treatments (DDAVP or blood products for surgical procedures), they are forced to change their self-image and to renounce to some activities for fear of bleeding. Thus, it can be easily understood that the importance of changing the approach of detection of VWD type 1 is relevant (Sadler, 2003).

1.2.2 Type 3

Type 3 VWD is a pathological disorder characterized by recessive inheritance of two null mutations that lead to extremely low or undetectable levels of von Willebrand factor and to moderate deficiency of FVIII. Mutations are distributed throughout the VWF gene and most are unique to the family which they were first identified. Usually the levels are very low (1-9 U/dL). Patients with type 3 VWD are therefore subjected to lifelong severe bleeding (Lillicrap, 2007).

1.2.3 Type 2A

Type 2A VWD is a pathological disorder, which shows qualitative abnormalities of VWF, due to a gene mutations. The VWF-dependent platelet adhesion is decreased because the proportion of large VWF multimers is decreased. Levels of VWF:Ag and FVIII may be normal or modestly decreased. The main feature of type 2A VWD is the absence of high and intermediate molecular weight multimers essential for coagulation and clinical trials show a low VWF:RCo to VWF:Ag ratio (<0.6) and abnormal RIPA. Mutations are located in the A2 domain of VWF and they can produce two different pathologies. Precisely, type 2AI VWD is characterized by the largest multimers retained in the cells of synthesis whereas in type 2AII VWD multimers are correctly synthesized and secreted into plasma, but then they are prematurely cleaved by the metalloproteinase ADAMTS-13. The inheritance of type 2A VWD is dominant with high penetration (Lillicrap, 2007).

1.2.4 Type 2B

Type 2B VWD is a pathological disorder characterized by an abnormal structure of the binding site for platelet glycoprotein α IIb caused by missenses mutations. The defect produces an increase in affinity of large multimers to platelets in the blood stream. Circulating platelets also are coated with mutant VWF, which may prevent the platelets from adhering at the site of the injury. Therefore, von

Willebrand factor is rapidly consumed. The inheritance of type 2B VWD is dominant with high penetration (Lillicrap, 2007).

1.2.5 Type 2N

Type 2N VWD is a pathological disorder characterized by low level of FVIII in the blood stream, caused by recessive mutations in the D'-D3 domains, thus the binding ability of VWF to FVIII is strongly reduced. The FVIII is a fundamental coagulation factor, known as antihemophilic factor. High levels of FVIII are linked with an increased risk of deep thrombosis and pulmonary embolism. FVIII is a glycoprotein released in the blood stream by vascular endothelial channels and sinusoidal liver cells. In the blood stream FVIII is linked with VWF, and together form an important coagulation complex (Lillicrap, 2007).

1.2.6 Type Vicenza

Type Vicenza VWD is a variant with plasma and platelet VWF level discrepancies and usually large VWF multimers. Its diagnosis is not easy due to the heterogeneous phenotype. The identification criteria consider the contribution of platelet VWF by comparison with plasma values. Clinical tests point out low plasma VWF but normal platelet VWF content and this can be a first useful indicator for its identification (Zieger, 1997). This type of VWD shows low or very low plasma VWF levels and the presence of ultra-large VWF meaning that the metalloproteinase ADAMTS-13 is not able to cut the multimers. Plasma VWF levels are the outcome of the synthesis and release of VWF from endothelial cells and its clearance from circulation. Most type of Vicenza patients have a normal platelet VWF content, which suggests a normal VWF synthesis. Moreover, after the DDAVP treatment the VWF level increases and this is an evidence of a normal acute release of VWF from both endothelial cells and Weibel Palade bodies. Thus, the low plasma level of VWF in type Vicenza VWD can be explained only by abnormalities occurring after the release of VWF. Indeed, type Vicenza VWD is characterized by an increased VWF clearance. This is an explanation of the VWD phenotype and the evidence that an increased clearance of VWF may be one of the cause of VWD. The most important step to follow for the identification of type Vicenza VWD are: at first evaluating the discrepancy between plasma and platelet VWF, then demonstrating a shorter VWF survival and the presence of ultralarge VWF multimers, in the end the identification of type Vicenza mutations (Casonato, 2006).

1.3 Diagnosis and detection of VWD

The diagnosis and the detection of the VWD types take into consideration three components: a personal history of excessive bleeding, a laboratory evaluation that discovers a quantitative or qualitative defect in VWF and the presence of a family history of excessive bleeding. In order to detect VWD the most common clinical trials are: the VWF:Ag to measure the number of VWF antigens, VWF:CB to evaluate the ability of VWF of binding with collagen, VWF:FVIII to evaluate the ability of VWF of binding with FVIII, RIPA (ristocetin induced platelet aggregation) to analyse the functionality of platelets and not only of the VWF and VWF:RCo, which is a functional test able to reflect its ability to aggregate normal platelets in the presence of ristocetin. Ristocetin is an antibiotic, that allows to link the VWF with the receptor glycoprotein α Ib on the surface of platelets. The combination of quantitative and functional tests enables the classification of the patients into VWD type (1, 2, 3) and subtype (2A, 2B, 2N, 2M, Vicenza). An important and quite invasive clinical test is the DDAVP, where desmopressin is administered subcutaneously at a dose of 40 mg/kg to patients and blood samples are collected before and after 15, 30, 60, 120, 180, 240, 480 and 24 h. DDAVP induces an acute release of VWF stored in the Weibel Palade bodies of the endothelial cells, followed by proteolysis of the UL multimers and VWF clearance. The VWF:Ag and VWF:CB after DDAVP are analysed by mean of a pharmacokinetic model describing the kinetic of variation in concentration (Casonato, 2006). Indeed, kinetically, the levels of VWF in plasma and patterns of HMW and LMW multimers depend on three determinants: i) the amount and rate of VWF release; ii) ADAMTS-13 proteolytic activity; and iii) VWF clearance from plasma. The DDAVP results elaborated by means of a pharmacokinetic model ensure the estimation of the amount and rate of VWF release, its clearance and half-life. Moreover, it enables to characterize the VWF kinetics of normal individuals with the 0 and non-0 blood group and for some types of VWD, even if the method cannot quantify the proteolytic activity of ADAMTS-13.

The main issues linked with the use of DDAVP as mean of detection are the fact that it is a long (at least 24 h), hazardous and invasive non-routine test, which needs to be carried out on the subjects to achieve a statistically satisfactory estimation of the individual metabolic parameters. However, pharmacokinetic models are now available and may be exploited to help in the diagnosis of VWD. Therefore, alternative less invasive and more easily implementable tests than DDAVP are recently receiving more attention in scientific community for model-based diagnostic purposes. In particular, these tests are the activated conformational state test (Groot, 2009) which helps the evaluation of VWF functional activity, the propeptide test (VWFpp) (Casonato, 2011) to quantify the VWF elimination from the blood stream and the interplatelet von Willebrand factor (Sweeney, 1992) to quantify the amount of VWF synthesized in the endothelial cells.

1.3.1 Clinical tests

In this section, the aim is to define shortly the standard procedure for the clinical tests that have been considered in our study. Clinical tests were carried out by the team of Prof. Alessandra Casonato at the Hospital of Padova in Italy. The study considers 20 VWD patients (types Vicenza, 2B) and 42 normal subjects, which were treated in accordance with the Helsinki Declaration, after obtaining their written informed consent and the Hospital of Padova approval of the ethical board.

1.3.1.1 VWF antigen (VWF:Ag)

Platelet VWF:Ag was measured with a home-made ELISA method, using washed platelets, adjusted to 1 million and lysed with Triton X-100. A pool of normal washed platelets was used to construct the reference curve. The results are given in U/dL, taking the first reference curve dilution as 100 U/dL.

1.3.1.2 VWF collagen binding (VWF:CB)

VWF:CB is assessed by ELISA using type III collagen (Sigma, Milan, Italy) diluted in acetic acid. The results are given in U/dL, taking the first reference curve dilution as 100 U/dL.

1.3.1.3 VWF propeptide (VWFpp)

Von Willebrand propeptide (VWFpp) was measured using a home-made ELISA method. Briefly, diluted reference and patient plasma samples were added to microwells on microtitration plates coated with a monoclonal antibody specific for VWFpp (CLB-Pro 35, Sanguin, The Netherlands); and bound VWFpp was assessed with a second anti-VWFpp HRP-labelled monoclonal antibody (M193904, Sanguin). The results are given in U/dL, taking the first reference curve dilution as 100 U/dL.

1.3.1.4 VWF:R and VWFpp ratio

VWFpp is used to assess the survival of VWF. It is secreted by endothelial cells as a dimer, in a ratio of 1:1 with mature VWF. While mature polymerized VWF survives for around 10-20 hours, VWFpp has a half-life of just 1-2 hours. No pathological conditions or mutations are known to affect the survival of VWFpp, while a number of VWF mutations are known to affect the half-life of mature VWF. That is why the VWFpp/VWF:Ag ratio (VWFpp ratio) gives an indirect measure of VWF survival. A reduced VWF half-life coincides with an increase in the VWFpp ratio: the higher the VWFpp ratio, the shorter the survival of VWF.

VWF:CB measures the capacity of VWF to bind to extravascular collagen. This binding relies on the integrity of the collagen binding domain of VWF, as well as the presence of large VWF multimers (the multimeric components best able to bind collagen). A lower VWF:CB/VWF:Ag ratio (VWF:CB ratio) is suggestive of a reduction in, or disappearance of large VWF multimers or, less frequently,

of an altered collagen binding domain of VWF. A greater reduction in the VWF:CB ratio coincides with a more pronounced shortage of large VWF multimers.

1.3.1.5 DDAVP

DDAVP (1-desamino-8-D-arginine vasopressin; Emosint, Sclavo, Italy) was administered subcutaneously at a dose of $0.3 \mu\text{g kg}^{-1}$. Blood samples were collected before and 15, 30, 60, 120, 180, 240, 480 min and 24 h after administering DDAVP. The time courses of the VWF:Ag and VWF:CB plasma concentrations after the DDAVP challenge were analysed using the SGM mathematical model that is described in section 2.1. In the study, this clinical trial is fundamental to calculate the values of the PK parameters for each VWD category and healthy subjects.

The clinical data derived from the execution of these medical assays have been fundamental for the development of the work here presented.

1.4 Pharmacokinetic model of VWD

Pharmacokinetic models are extensively used nowadays in preliminary drug development, in preclinical studies for the formulation of new therapies and in clinical diagnosis. PK models have been developed to conceptualize, in simple terms, the processes that take place between an organism and a chemical substance. Moreover, prediction by PK simulation can reduce the in vivo experiments, anticipate the response of patients to new drugs increasing both efficacy and safety. Again, knowing the desired response, model based design of experiments techniques can be applied to optimise clinical studies in different stages starting from diagnosis (Galvanin, 2013).

The PK model taken as reference in this research work is a simplified version of the model proposed by Galvanin et al. (2017), that is a two compartments model designed to analyse the time courses of plasma VWF:Ag and VWF:CB levels after DDAVP clinical trials. From now on, this model will be referred to as SGM (simplified Galvanin model).

As reported in figure 1.3, the SGM is able to characterize the mechanisms of VWF release, proteolysis, clearance and the multimers distribution of VWF in the plasma. The model is described by a system of differential and algebraic equations where each subject is characterized using three main PK constants, namely the VWF release rate k_0 [h^{-1}], the proteolysis rate k_1 [h^{-1}] and the elimination rate k_e [h^{-1}], which is assumed to be the same for both the UL+HMW multimers and the LMW multimers. The amount of released VWF is represented by parameter D [U/dL]. It is important to notice that, for a given subject, parameter k_0 quantifies the rate of release, while D quantifies the amount of VWF released from the endothelial cells.

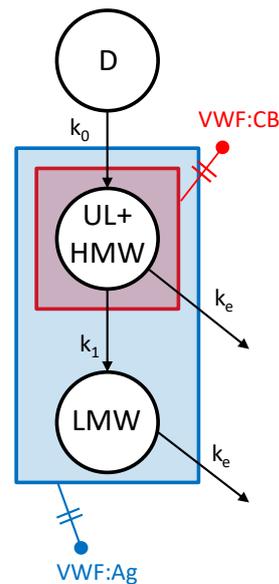


Figure 1.3. SGM model structure; VWF:Ag and VWF:CB measurements are identified by dashed boxes; D = release of UL+HMW multimers after DDAVP administration (UL = ultra-large; HMW = high-molecular-weight; LMW = low-molecular-weight).

The physiological assumptions at the core of the model are:

1. UL, HMW and LMW multimers are present in the basal state and/or after DDAVP;
2. UL and HMW multimers can be cleaved to form LMW multimers;
3. The VWF:Ag measurements enable to assess the quantities of UL+HMW+LMW multimers;
4. The VWF:CB measures the UL+HMW multimers.

The use of the PK model allows to estimate the model parameters through DDAVP measurements which are paradigmatic for each category of disease.

The mathematical structure that define the model and the parameter estimation procedure are described in section 2.1.

1.5 Objectives

As mentioned in §1.3, the SGM of VWD requires the execution of a long-lasting (24 h) DDAVP test to estimate the model kinetic parameters. Therefore, the overall objective of the project is to optimally design a minimal set of basal clinical trials for the identification of the PK model of VWD, in order to decrease the time and effort required for disease characterization and diagnosis. The idea is to reduce the execution time of the DDAVP modifying the SGM by introducing explicit correlations, which relate the PK parameters to basal clinical quantities.

A more ambitious goal is to eliminate completely the execution of the DDAVP clinical trial. This challenge is tackled substituting in the SGM basal state correlations for all the model kinetic parameters.

Suitable relations between the available experimental data are investigated to reach the target. Furthermore, the applicability of regression models needs to be verified for building the mathematical structure of the correlations.

Once the basal state equations are defined, they are substituted in the SGM and the DDAVP can be shortened or even avoided.

The statistical tools required for achieving the main goals of the work of thesis are presented in the following chapter.

Chapter 2

Identification and modification of VWD pharmacokinetic model

In this chapter, the description of the SGM model of VWD and the methodology used to achieve the research goals are described. Moreover, a brief explanation of the procedure applied for model modification and identification is given.

2.1 The SGM mathematical structure

The SGM is described as a set of differential and algebraic equations. The two differential equations are written as:

$$\frac{dx^{UL+HMW}}{dt} = k_0 D e^{-k_0(t-t_{max})} - k_1(x^{UL+HMW} - x_b^{UL+HMW}) - k_e(x^{UL+HMW} - x_b^{UL+HMW}) \quad (2.1)$$

$$\frac{dx^{LMW}}{dt} = k_1(x^{UL+HMW} - x_b^{UL+HMW}) - k_e(x^{LMW} - x_b^{LMW}) \quad (2.2)$$

where: x^{UL+HMW} and x^{LMW} are the number of UL+HMW and LMW multimer units [U] contained in the plasma; the subscript b refers to the basal state; t is the time; and t_{max} is the time at which the release profile peaks. In the PK model k_0 represents the kinetics of VWF release from endothelial cells; k_1 the proteolytic conversion of large and ultra-large VWF multimers into LMW multimers and k_e represents the clearance of VWF from the circulation. The amount of VWF released after DDAVP is measured by the parameter D , and the release time is represented by parameter t_{max} .

The antigen concentration y^{AG} [U/dL] and collagen binding concentration y^{CB} [U/dL] are defined as:

$$y^{AG} = \frac{x^{UL+HMW} + x^{LMW}}{V_d} \quad (2.3)$$

$$y^{CB} = \frac{x^{UL+HMW}}{V_d} \quad (2.4)$$

where: $V_d = 40 \text{ mL/kg}_{\text{bw}}$ is the approximate distribution volume. Basal conditions are assumed at $t(0)$, i.e. $x(0) = [x_b^{UL+HMW} \quad x_b^{LMW}] = [y_b^{CB}V_d \quad y_b^{AG}V_d - y_b^{CB}V_d]$. As suggested by Galvanin et al. (2014), a correction was introduced in the definition of the collagen binding measurements to account for the different affinity of multimers to collagen observed in different types of VWD:

$$y^{CB'} = k y^{CB} \frac{y_b^{AG}}{y_b^{CB}} \quad (2.5)$$

where k is the correction factor.

2.1.1 Parameters estimation procedure

The set of parameters to be estimated from the available measurements is $\theta = [k_0 \quad k_1 \quad k_e \quad D \quad k \quad y_b^{CB} \quad t_{max}]$. Their estimation is based on the maximum likelihood principle and carried out using the commercial software gPROMS[®]. Measurements are assumed to be normally distributed with a standard deviation of 2 U/dL. The parameter set θ is determined for each individual by iteratively solving an optimization problem. A crucial step in the parameter estimation exercise is setting the value of the first guess θ_0 for θ . In this study, θ_0 was obtained by carrying out a preliminary parameter estimation for each group of subjects (healthy, VWD type 2B, and VWD type Vicenza) using fictitious concentration profiles corresponding to the average profiles of all the subjects belonging to a given group. Then the parameter estimation routine was initialized with θ_0 , and the set of parameters was estimated for each individual subject based on his/her own concentration profiles.

A two-step iterative procedure was used to estimate the parameters, basing on each subject's VWF:Ag and VWF:CB readings:

Step 0 – all parameters ($k_0 \quad k_1 \quad k_e \quad D \quad k \quad y_b^{CB} \quad t_{max}$) are left free to vary starting from the initialization value θ_0 ;

Step 1 – [$D \quad k \quad y_b^{CB} \quad t_{max}$] are set at the value identified in the previous step, and the kinetic parameters [$k_0 \quad k_1 \quad k_e$] are estimated;

Step 2 – [$k_0 \quad k_1 \quad k_e \quad D$] are set at the value used in the previous step, and the correction parameters [$k \quad y_b^{CB}$] and t_{max} are estimated;

Steps 1 and *2* are repeated until the estimates do not vary significantly (i.e. until the difference between the estimates is lower than 0.1% for each parameter).

Parameter estimation results are given in the appendix A for each subject in terms of estimated values and related statistics on parameter estimation. Knowing the values of the PK parameters of each subject considered in the study is fundamental for the application of RSM technique necessary for model modification.

2.2 Methodology

2.2.1 Model-based design of experiments (MBD_{oE})

Models can present different strength and weaknesses, degrees of complexity and descriptive capabilities. Complex models might be capable of realizing low residuals when they are used to fit the experimental data, but too high complexity may result in a very difficult identification of the parameters. However, the identification of the model requires the execution of experiments, which may involve the employment of costly facilities, resources and time. It is therefore of great importance planning the experiments carefully, taking into account the specific target, which, in this case, is the identification of the modified PK model of VWD through the improvement of the estimates of its parameters. Many researchers devoted their efforts to develop advanced MBD_{oE} techniques for both model design (MD) and parameters precision (PP) (Box & Lucas, 1959) (Espie & Macchietto, 1989). These theories allow for the identification of the best experiment campaign also in complex nonlinear dynamic systems through the numerical maximization/minimization of properly defined objective functions. If the target is to improve the precision of the parameters, then the objective function to minimize is a certain measure of the predicted covariance matrix associated to the parameter estimates \mathbf{V}_θ . The versatile mathematical framework in which these theories were developed made possible their application in different branches of science: automotive, nuclear physics, medicine, kinetics and many others (Bard, 1974). The following sections are dedicated to present the mathematical tools that have been applied in this thesis.

2.2.1.1 Information content analysis

Validation and identification of the simplified model of Galvanin et al. (2017) was conducted in a previous research work. The model has been found capable of describing the proteolysis of VWF in the individual subjects provided the correct PK parameters.

In order to optimally design a minimum set of clinical tests for the characterization of the disease, the information content of the system needs to be maximized. The application of DoE techniques aims at

minimizing the variances of the experimental measurements, which are then related to the uncertainty on the estimated parameters. This goal can be achieved defining the optimal experimental sets and measurement times leading to the maximum information content derivable from the experiment (Fedorov, 2014).

Considering the SGM (2017) sensitivity analysis has been performed and Fisher information matrix has been calculated to evaluate the time required by the parameters to achieve the maximum information content from the DDAVP clinical trial. This step is fundamental in order to understand which parameters are more critical and which direction should be taken for optimally modify the pharmacokinetic model.

Sensitivity analysis is carried out assigning a small perturbation (1%) to the PK parameters acting on one factor at a time and observing how it affects the predictive capability of the model itself. Sensitivity analysis was considered for both the model responses: antigen concentration y_{Ag} and collagen-binding concentration y_{CB} . In mathematical terms this can be written as:

$$q_i^{Ag} = \frac{y_{Ag}(\theta_i') - y_{Ag}(\theta_i)}{\theta_i' - \theta_i} \quad i=1, \dots, N_\theta \quad (2.5)$$

$$q_i^{CB} = \frac{y_{CB}(\theta_i') - y_{CB}(\theta_i)}{\theta_i' - \theta_i} \quad i=1, \dots, N_\theta \quad (2.6)$$

Where θ_i and θ_i' represent the original and perturbed sets of parameters, respectively and N_θ is the number of model parameters.

Sensitivity analysis gives only a qualitative idea of the information, whereas the information is expressed quantitatively by the Fisher information matrix \mathbf{H} . Considering the maximum likelihood estimate presented in §2.4.3.1, if Φ is defined as the logarithm of the likelihood function, the Hessian:

$$\mathbf{H} = \frac{\partial^2 \Phi}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} = \frac{\partial^2 \ln(L(\hat{\boldsymbol{\theta}}))}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \quad (2.7)$$

is also called Fisher information matrix (FIM) and it quantifies the information carried by measurable random variable \mathbf{z} about non-measurable unknown parameters $\boldsymbol{\theta}$. Due to the fact that this matrix is the Hessian of a log-likelihood function, also the approximation $\mathbf{V}_\theta \cong \mathbf{H}^{-1}$ holds (§2.4.3.2). Intuitively, the covariance matrix \mathbf{V}_θ and the Fisher matrix \mathbf{H} represent two sides of the same coin. The higher the information carried by the measurements, the lower the uncertainty associated to the estimated parameters. Assumed that a model is given to describe a certain physical phenomenon and that some experiments have been performed to obtain a first rough estimation of the parameters, the first estimate is $\boldsymbol{\theta}^0$ and the associated covariance matrix is \mathbf{V}_{θ^0} . The goal is to improve the estimate reducing the elements of the covariance matrix associated to the parameters performing new

experiments. It is possible to quantify approximately in advance the posterior covariance matrix \mathbf{V}_θ after the conduction of N_{exp} experiments as:

$$\mathbf{V}_\theta \cong \left[\mathbf{V}_{\theta^0}^{-1} + \sum_{i=1}^{N_{exp}} \mathbf{H}_i \right]^{-1} \quad (2.8)$$

Where \mathbf{H}_i represents the information matrix associated to the i -th experiment in a hypothetical campaign of N_{exp} experiments. Under certain conditions, it is possible to use an approximate form of the Fisher matrix. In particular, it is possible to refer to a log-likelihood function considering totally uncorrelated measured variables:

$$\Phi = \ln(L(\hat{\theta})) = \frac{1}{2} \sum_{i=1}^{N_{exp}} \sum_{j=1}^{N_m} \left[\log(2\pi\sigma_{ij}^2) + \left(\frac{\hat{y}_{ij} - y_{ij}}{\sigma_{ij}} \right)^2 \right] \quad (2.9)$$

The kl -th element of the Fisher information matrix is also defined as the kl -th element of the Hessian matrix associated to function Φ :

$$[\mathbf{H}]_{kl} = \left[\frac{\partial^2 \ln(\Phi)}{\partial \theta_k \partial \theta_l} \right]_{kl} = \sum_{i=1}^{N_{exp}} \sum_{j=1}^{N_m} \left[\frac{1}{\sigma_{ij}^2} \left(\frac{\partial \hat{y}_{ij}}{\partial \theta_k} \frac{\partial \hat{y}_{ij}}{\partial \theta_l} \right) + \frac{1}{\sigma_{ij}^2} \left(\hat{y}_{ij} - y_{ij} \right) \frac{\partial^2 \hat{y}_{ij}}{\partial \theta_k \partial \theta_l} \right] \quad (2.10)$$

If residuals are small, it is acceptable to write the following approximation:

$$[\mathbf{H}]_{kl} \cong \sum_{i=1}^{N_{exp}} \sum_{j=1}^{N_m} \left[\frac{1}{\sigma_{ij}^2} \left(\frac{\partial \hat{y}_{ij}}{\partial \theta_k} \frac{\partial \hat{y}_{ij}}{\partial \theta_l} \right) \right] \quad (2.11)$$

The term $\frac{\partial \hat{y}_{ij}}{\partial \theta_k}$ is called sensitivity of the j -th output variable with respect to the k -th parameter in the conditions investigated in the i -th experiment. Throughout this work, notation (11) for the Fisher matrix is used to plot the information in the experimental design space to visualize the most informative experimental conditions (Bard, 1974).

The trace of the Fisher matrix has been calculated and is used to represent the overall information content for a hypothetical discrete sampling times (t_{sp}) and it is adopted as suitable scalar measure of the information:

$$I_d = \sum_{t_{sp}} \text{tr}[\mathbf{H}_\theta] \quad (2.12)$$

2.2.2 Parameter estimation problem

When a model has already been selected among a set of candidates to describe a system, its identification reduces to the estimation of its parameters. In this paragraph, the model identification problem is presented taking into consideration the uncertainty intrinsically linked to the

measurements and also to the description of the statistical instrument necessary to assess the quality of the results (Bard, 1974).

Be $\hat{\mathbf{y}}_i = \mathbf{f}(\mathbf{x}, \mathbf{u}_i, \boldsymbol{\theta})$ the vector of output variables predicted by the model in the experimental conditions adopted in the i -th experiment, \mathbf{x} is the vector of the state variables, \mathbf{u}_i the input variables that can be modified and $\boldsymbol{\theta}$ the vector of the model parameters, the quantity $\rho_{ij}(\boldsymbol{\theta})$, named as residual, represents the difference between measured and predicted value for the j -th output variable in the i -th experiment:

$$\rho_{ij}(\boldsymbol{\theta}) = y_{ij} - \hat{y}_{ij}(\boldsymbol{\theta}) \quad (2.13)$$

The parameter estimation problem is then recast in terms of finding the best set of parameters $\boldsymbol{\theta}$ that minimizes a certain objective function Φ that depends on the quantity $\rho_{ij}(\boldsymbol{\theta})$. This function could be defined as the sum of the squared residuals, but this approach does not take into account the uncertainty intrinsically associated to the measurements. The conditions under which a model is identified are never quite repeatable because of the random nature and the limited accuracy of any measurement technique. These disturbances are as much part of the physical reality as are the quantities appearing in the model. A model cannot be called complete if it does not take into account the casual nature of the measurements. Therefore, the appropriate description of random events is made through the concept of probability. In a rigorous mathematical description, the complete characterization of the behaviour of a random variable x is given by the definition of an associated probability density function PDF which associates a probability of realisation to any possible value of the variable. One of the most popular PDF is represented by:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (2.14)$$

Which is the univariate normal distribution with mean μ and standard deviation σ . The success of the normal distribution is not only due to its easily treatable mathematical structure, but also to the fact that it has been discovered to describe closely the errors associated to many measurements in nature (Bard, 1974).

2.2.2.1 *The maximum likelihood estimate*

Assuming that the difference between the measured and the true values of the quantities appearing in the model as output variables $y_{ij} - y_{ij}^*$ are normally distributed random variables with zero mean and a certain standard deviation (SDV) σ_{ij} and that the model used to fit the experimental data is correct, then a true value does exist for the set of parameters $\boldsymbol{\theta}^*$ such that the model prediction is exact. Thus,

for that particular value of the parameters, the residuals $\rho_{ij}(\boldsymbol{\theta}^*)$ follow the same distribution of the measurement errors $y_{ij} - y_{ij}^*$, indeed:

$$\rho_{ij}(\boldsymbol{\theta}^*) = y_{ij} - \hat{y}_{ij}(\boldsymbol{\theta}^*) = y_{ij} - y_{ij}^* \quad (2.15)$$

Consider now the joint probability density function of the residuals $\rho_{ij}(\boldsymbol{\theta})$, assumed as completely uncorrelated, normally distributed random variables with zero mean and standard deviation equals to the SDV of the associated measurement σ_{ij} :

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N_{exp}} \prod_{j=1}^{N_m} \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_{ij}}\right)^2} \quad (2.16)$$

The joint PDF of the residuals is also called likelihood function. The parameter estimation problem, can be reformulated in terms of finding the values of the parameters $\boldsymbol{\theta}$ which maximizes the objective function $L(\boldsymbol{\theta})$, causing the final residuals obtained after maximization to be distributed like the corresponding measurement error (Bard, 1974):

$$\max_{\boldsymbol{\theta}}\{L(\boldsymbol{\theta})\} = \max_{\boldsymbol{\theta}} \left\{ \prod_{i=1}^{N_{exp}} \prod_{j=1}^{N_m} \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_{ij}}\right)^2} \right\} \quad (2.17)$$

This definition is implemented in gPROMS to estimate the PK parameters in the modified model.

Assuming to use an exact structural model, it is rigorously acceptable to declare that residuals and measurements errors follow the same distribution. However, in a quasi-exact model this simplification is not cause of much harm and the non-perfect structure of the model is usually detected through a posterior analysis (i.e χ^2 -test). Moreover, the number of measurements available will be always limited and might not be sufficient for a reliable estimate of the model parameters.

A number of tools, described below, are given to the modelers in order to asses the quality of the fitting and of the estimates.

2.2.2.2 The covariance matrix

A certain objective function Φ , has been chosen to be maximized/minimized in order to estimate the set of model parameters $\boldsymbol{\theta}$. It is not possible to compute a value $\hat{\boldsymbol{\theta}}$ and state that the estimate obtained represents the “true” values of the parameters; in fact the computed value obtained depends on the measured values \mathbf{y}_i ($i = 1, 2, 3, \dots, N_{exp}$), which are affected by uncertainty. This paragraph is made to explain the mathematical tool necessary to assess how the uncertainty associated to the measured

values impacts the confidence we can assign on the estimated non-measurable parameters. Assume that $\boldsymbol{\theta}$ is an N_θ vector of model parameters and that a certain value $\hat{\boldsymbol{\theta}}$ has been computed maximizing/minimizing an objective function Φ .

A new column vector \mathbf{z} with dimensions $N_{exp}N_m$ which contains all the vectors \mathbf{y}_i ($i = 1, 2, 3, \dots, N_{exp}$) is defined:

$$\mathbf{z} = \left[y_{11}, \dots, y_{1N_m}, \dots, \dots, y_{N_{exp}1}, \dots, y_{N_{exp}N_m} \right]^T \quad (2.18)$$

Also two other column vectors $\delta\hat{\boldsymbol{\theta}}$ and $\delta\mathbf{z}$ are defined. In particular, $\delta\hat{\boldsymbol{\theta}}$, with dimension N_θ , represents a shift of the estimated value $\hat{\boldsymbol{\theta}}$ for the parameters derived from a variation $\delta\mathbf{z}$ in the measured values for the output variables. If $\hat{\boldsymbol{\theta}}$ has been computed maximizing/minimizing the objective function Φ , the following condition is satisfied:

$$\frac{\partial\Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial\boldsymbol{\theta}} = \mathbf{0} \quad (2.19)$$

Where the left-hand term represents the column vector of dimension N_θ whose elements represent the partial derivatives of the objective function Φ with respect to the parameters. If the function Φ is continuous, a small variation in the measured values $\delta\mathbf{z}$ results in a small shift of the computed value $\hat{\boldsymbol{\theta}}$ in the space of the parameters:

$$\frac{\partial\Phi(\hat{\boldsymbol{\theta}} + \delta\hat{\boldsymbol{\theta}}, \mathbf{z} + \delta\mathbf{z})}{\partial\boldsymbol{\theta}} = \mathbf{0} \quad (2.20)$$

Expanding the condition to the first term of Taylor expansion it is possible to quantify approximately the variation of the estimated value for the parameters shift $\delta\hat{\boldsymbol{\theta}}$.

$$\frac{\partial\Phi(\hat{\boldsymbol{\theta}} + \delta\hat{\boldsymbol{\theta}}, \mathbf{z} + \delta\mathbf{z})}{\partial\boldsymbol{\theta}} \cong \frac{\partial\Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial\boldsymbol{\theta}} + \frac{\partial^2\Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}} \delta\hat{\boldsymbol{\theta}} + \frac{\partial^2\Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial\boldsymbol{\theta}\partial\mathbf{z}} \delta\mathbf{z} \quad (2.21)$$

Note that the first term of the expansion is equal to $\mathbf{0}$ because it has been supposed that $\hat{\boldsymbol{\theta}}$ represents an extremum point of the objective function. Also notice that:

$$\frac{\partial^2\Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}} = \mathbf{H} \quad (2.22)$$

Represents the symmetric Hessian matrix of function Φ evaluated with respect to the parameters whose kl -th element is:

$$[\mathbf{H}]_{kl} = \frac{\partial^2\Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial\theta_k\partial\theta_l} \quad (2.23)$$

From the Taylor series the term $\delta \hat{\boldsymbol{\theta}}$ is isolated:

$$\frac{\partial^2 \Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \delta \hat{\boldsymbol{\theta}} + \frac{\partial^2 \Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial \boldsymbol{\theta} \partial \mathbf{z}} \delta \mathbf{z} \cong \mathbf{0} \quad (2.24)$$

$$\delta \hat{\boldsymbol{\theta}} \cong -\mathbf{H}^{-1} \left(\frac{\partial^2 \Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial \boldsymbol{\theta} \partial \mathbf{z}} \right) \partial \mathbf{z} \quad (2.25)$$

The covariance matrix associated to the estimates is defined as the expected value of the squared deviation of the parameters from their expected value $E(\boldsymbol{\theta})$:

$$\mathbf{V}_{\boldsymbol{\theta}} = E\{[\boldsymbol{\theta} - E(\boldsymbol{\theta})][\boldsymbol{\theta} - E(\boldsymbol{\theta})]^T\} \quad (2.26)$$

If it is assumed that $E(\boldsymbol{\theta}) = \hat{\boldsymbol{\theta}}$ then the covariance matrix associated to the parameter estimates, approximated to the first term of the Taylor expansion, is evaluated as follows:

$$\mathbf{V}_{\boldsymbol{\theta}} \cong E \left[\left(-\mathbf{H}^{-1} \left(\frac{\partial^2 \Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial \boldsymbol{\theta} \partial \mathbf{z}} \right) \partial \mathbf{z} \right) \left(-\mathbf{H}^{-1} \left(\frac{\partial^2 \Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial \boldsymbol{\theta} \partial \mathbf{z}} \right) \partial \mathbf{z} \right)^T \right] \quad (2.27)$$

$$\mathbf{V}_{\boldsymbol{\theta}} \cong E \left[\mathbf{H}^{-1} \left(\frac{\partial^2 \Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial \boldsymbol{\theta} \partial \mathbf{z}} \right) \partial \mathbf{z} \partial \mathbf{z}^T \left(\frac{\partial^2 \Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial \boldsymbol{\theta} \partial \mathbf{z}} \right)^T \mathbf{H}^{-1} \right] \quad (2.28)$$

Notice that the only term containing random variable is $\partial \mathbf{z} \partial \mathbf{z}^T$, which represents the $N_{exp} N_m \times N_{exp} N_m$ covariance matrix associated to the measurement \mathbf{V}_z . It is therefore possible to rewrite explicitly the covariance matrix of the estimates:

$$\mathbf{V}_{\boldsymbol{\theta}} \cong \mathbf{H}^{-1} \left(\frac{\partial^2 \Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial \boldsymbol{\theta} \partial \mathbf{z}} \right) \mathbf{V}_z \left(\frac{\partial^2 \Phi(\hat{\boldsymbol{\theta}}, \mathbf{z})}{\partial \boldsymbol{\theta} \partial \mathbf{z}} \right)^T \mathbf{H}^{-1} \quad (2.29)$$

The formula applies for every choice of the objective function, however, for a specific class of functions Φ including the sum of squared residuals and the natural logarithm of the likelihood function, it can be demonstrated that also the following approximation holds:

$$\mathbf{V}_{\boldsymbol{\theta}} \cong \mathbf{H}^{-1} \quad (2.30)$$

The quality of the above approximation improves as the variance of the measurements decreases and the fitting of the model gets better (Bard, 1974).

2.2.2.3 The t -test

The comparison of the variances associated to the estimated vector $\hat{\theta}$ already gives good information about the parameters that require more attention and the possible critical structural weaknesses of the model. However, to assess the statistical quality of the parameters, it is necessary to compare the value of each parameter estimated within its confidence region. Precisely, it is important to define the dimension of the confidence region assigned to the parameter with respect to the absolute value of the parameter itself. A parameter estimation problem involving N_θ parameters is solved assuming a dataset of $N_{exp} \times N_m$ measurements. In this thesis, to assess the statistical value of the estimates, a one-tailed t -test with 95% of significance is performed comparing the t -value of each estimated parameter $\hat{\theta}_i$ with the reference t -value of a Student distribution with degree of freedom $N_{exp}N_m - N_\theta$:

$$\frac{\hat{\theta}_i}{t_{0.975}(N_{exp}N_m - N_\theta)\sqrt{V_{\theta,ii}}} > t_{0.95}(N_{exp}N_m - N_\theta) \quad \forall i=1, \dots, N_\theta \quad (2.31)$$

Where the t -value appearing in the bottom part of the left-hand term is evaluated for a Student distribution with degree of freedom $N_{exp}N_m - N_\theta$ at a cumulated probability equals to 0.975 and the t -value of reference appearing in the right-hand side is evaluated at a cumulated probability of 0.95 to perform the one-tailed test with 95% of significance. The satisfaction of this condition is considered as a proof of good estimation of the parameters (Bard, 1974).

2.2.2.4 The Chi-squared for the goodness of fit evaluation

In a conventional parameter estimation problem, a proposed model is used to fit a set of data, indeed the model used might not reflect exactly the nature of the physical phenomenon. In this work, to detect a bad fitting a χ^2 -test on the residuals with 95% of significance is performed. The χ^2_{ref} depends on the number of degrees of freedom $N_{exp}N_m - N_\theta$ specific of each case. The test is important to understand if the residuals computed at the end of the parameter estimation problem can be justified by the measurements errors. Summing up $N_{exp}N_m - N_\theta$ squared random variables following the standard normal distribution, the result will be smaller than χ^2_{ref} with a probability of 95%. The reference value is compared to the squared weighted residuals obtained as solution of the parameter estimation problem:

$$\chi^2_{sample} = \sum_{i=1}^{N_{exp}} \sum_{j=1}^{N_m} \left[\frac{\rho_{ij}(\hat{\theta})}{\sigma_{ij}} \right]^2 \quad (2.32)$$

If the model is exact and the sample used for the fitting is sufficiently large, the values computed for $\hat{\theta}$ are expected to be very close to the “true” values, thus the model identified would be very close to the “true” model and the residuals would be consequence of the measurement errors only. If these errors are normally distributed and the values of the SDVs associated to the measurements are known precisely and not underestimated, then $\chi^2_{sample} \leq \chi^2_{ref}$ with a probability of 95% (Bard, 1974). If we are not sure about the measurement uncertainty and about the reliability of the model, and if it happens $\chi^2_{sample} > \chi^2_{ref}$, the errors can be interpreted in 5 different ways:

- the available experimental data are not sufficient to estimate correctly the parameters;
- the assumption of having measurements errors following a normal distribution with zero mean is wrong;
- the value of the SDVs associated to the measurements have been globally underestimated;
- the model is wrong;
- a combination of the 4 previous cases.

2.2.3 MBD_{oE} for parameter estimation

A model is available together with preliminary experimental data, therefore the solution of a parameter estimation problem leading to the computation of first set of parameters θ^0 is possible. The evaluation of the covariance matrix V_{θ^0} and the t -tests performed on the parameters allow evaluating whether a satisfactory estimation of the parameters may be achieved or some parameters are affected by strong correlation and very high variance. If the second case occurs, it is necessary to amend the unsatisfactory estimates performing new experiments. As presented in § 2.4.3.2, it is possible to quantify approximately the posterior covariance matrix V_{θ} resulting by the execution of a certain set of N_{exp} experiments through the evaluation of the FIM. By doing so, it is possible to design an experiment campaign with the aim of minimizing a certain measure of the posterior covariance matrix V_{θ} . In general, the covariance matrix of the estimates identifies a confidence ellipsoid in the N_{θ} -dimensional hyperspace. Improving the parameters estimates means reducing the size of this region of confidence choosing the proper scalar measure as target to minimize. Different meaningful scalar quantities can be chosen as objective function, but the most established and popular methods are:

- A-optimal: which consider the trace of V_{θ} as scalar function to be minimized. The trace of the covariance matrix associated to the parameter estimates quantifies the volume of the polyhedron circumscribing the confidence ellipsoid in the N_{θ} -dimensional space of the estimates;

- D-optimal: for which the determinant of the matrix \mathbf{V}_θ is chosen as objective function. The determinant of the covariance matrix quantifies the volume of the confidence ellipsoid;
- E-optimal: in which the largest eigenvalue of \mathbf{V}_θ is assumed as measure to minimize. The largest eigenvalue of the covariance matrix quantifies the length of the longest axis of the confidence ellipsoid.

2.2.4 Response surface methodology (RSM)

Response surface modelling (RSM) is a technique strictly related to the design of experiment, used to develop black-box models (regression models), searching for a correlation between inputs and outputs variables in systems where no information is given on what is happening inside. Considering the seminal work by Box-Behnken (1960), this approach is used to interpolate or approximate the information coming from experimental data, with the aim of defining the profile of the response in the experimental workspace. The objective consists into hypothesize an analytical form of the response surface, which manages to fit or approximate the experimental data reducing the distance between real and simulated response. In this way, the error of the model is reduced and the response can be estimated from the inlet variables. Mathematically, the response model surface is an approximating k-dimensional hyper-surface, acting in a space k+1 dimensional, made of k factors and the output function. In the study, we are interested in defining explicit correlations for the PK parameters k_0 , k_1 , k_e as a function of two inlet variables, which are represented by a combination of basal clinical trials. Therefore, the workspace we are dealing with is 3D, whereas the hypersurface is 2D. The advantage of the response surface modelling consists in the possibility of representing the combinations of input variables suitable to obtain the desired response with the lowest error. The analytical forms of the response surface models can be various; for instance, possible choices are:

1. Linear response surface model without interaction

$$\hat{f}(x) = b_0 + \sum_{j=1}^k b_j x_j \quad (2.33)$$

2. Linear response surface model with interactions

$$\hat{f}(x) = b_0 + \sum_{j=1}^k b_j x_j + \sum_{j=1}^{k-1} \sum_{j < r} b_{jr} x_j x_r \quad (2.34)$$

3. Quadratic response surface model

$$\hat{f}(x) = b_0 + \sum_{j=1}^k b_j x_j + \sum_{j=1}^k \sum_{r=1}^k b_{jr} x_j x_r \quad (2.35)$$

Based on Akaike index (Akaike, 1974) evaluations and knowing the physiology of the disease considered, linear response surface model with interactions appears to be the best candidate correlation for truly representing the real system. The Akaike index is a measure of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each other models. Hence, AIC (Akaike index criterion) provides a means for model selection (Akaike, 1974). The AIC value of the model is the following:

$$AIC=2k-2\ln(\hat{L}) \quad (2.36)$$

Where k is the number of model parameters, whereas \hat{L} is the maximum value of the likelihood function for the model.

In the fitting procedure, the values of the parameters are evaluated through the Least squared method, suitable for the overdetermined systems, that is, the number of parameters that needs to be estimated is lower than the available experimental points. The reliability of the response provided by the analytical form of the hypersurface increases with the increase in the amount of the experimental data available.

2.2.4.1 Goodness of fit

Response surface modelling approach has been applied in OriginPro[®] graphics and data analyser. Indeed, after data fitting, it is important to evaluate its goodness. A visual examination of the fitted curve displayed in the curve fitting should be the first step. Different goodness of fit measures can be used for both linear and nonlinear parametric fits and they are described as follow:

- Residuals

The residuals from a fitted model are defined as the difference between the response data and the fit to the response data at each predictor value. Assuming the model that fits the data is correct, the residuals approximate the random errors. Therefore, if the residuals appear to behave randomly, it suggests that the model fits the data well. However, if the residuals display a systematic pattern, it is a clear sign that the model fits the data poorly. Mathematically they are described as:

$$r_i = \hat{y}_i - y_i \quad (2.37)$$

- Goodness of fit statistics

The sum of squares due to error (SSE) shows the total deviation of the response values from the fit to the response values:

$$SSE = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \quad (2.38)$$

Where \hat{y}_i is the estimate by the regression model, while \bar{y} is the average of the observed data y_i . A value closer to 0 indicates a better fit.

R-square measures how successful the fit is in explaining the variation of the data. R-square is the square of the correlation between the response values and the predicted response values. This statistic is defined as the ratio of the sum of squares (SSE) of the regression and the total sum of squares (SST):

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \quad (2.39)$$

where SSR is the residual sum of squares:

$$SSR = \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (2.40)$$

and SST is the total sum of squares:

$$SST = \sum_{i=1}^N (y_i - \bar{y})^2 \quad (2.41)$$

R^2 can take only values between 0 and 1, a value closer to 1 indicates a better fit. If the number of fitted coefficients in the model increases, R^2 might increase although the fit may not improve. To avoid this situation, the \bar{R}^2 statistic can be used. It is possible to get negative R^2 for equations that do not contain a constant term.

\bar{R}^2 uses the R^2 statistic described above and adjust it based on the residual degrees of freedom. The residual degrees of freedom is defined as the number of response values n minus the number of fitted coefficients m estimated from the response values. Mathematically it is described as follow:

$$\bar{R}^2 = 1 - \frac{SSR/df_e}{SST/df_t} \quad (2.42)$$

Where df_e are the degrees of freedom of the estimate on the underlying population error variance, while df_t are the total degrees of freedom of the system considered.

The \bar{R}^2 statistic can take only values less or equal to 1, with a value closer to 1 indicating a better fit. Root mean squared error (RMSE) is also known as the fit standard error and the standard error of the regression. A RMSE value closer to 0 indicates a better fit.

Once suitable correlations are defined, data mining approach is applied to determine the coefficients of the surface for each category of disease and healthy subject in the most precise way.

2.4.6 Software

The entire RSM has been carried out in OriginPro[®] data analyser by OriginLab corporation. OriginPro[®] offers extended analysis tools for statistics 3D fitting, image processing and signal processing (2007). In particular, non-linear curve fitting can be performed with user-defined functions and this has been used to develop and test the regression model.

The parameter estimation and the information content analysis of the simplified model by Galvanin et al. and the model identification procedure and simulations of the modified model have been instead carried out in gPROMS[®] Model Builder environment 4.1.0. gPROMS[®] is an advanced modelling software by PSE in London. The software is a dynamic equation oriented simulator, which allows to solve robustly large scale DAEs systems (2004). One of the main characteristics of the software is that it possesses a powerful optimization and parameter estimation tools, which allows a trustworthy resolution of the parameter estimation problem with a great accuracy (2004).

2.3 Model modification and identification approach

In this paragraph, a summary of the main steps of the work is given to contextualize the described methodology in the development of the work of thesis. The overall steps are defined in the block-diagram reported in figure 2.1.

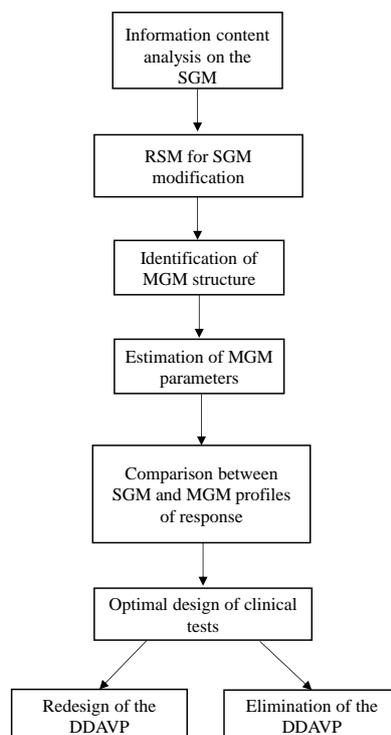


Figure 2.1. Summary of the overall steps of the project.

Information content analysis on the SGM is executed to discover the DDAVP execution time required for the identification of the PK parameters. Then, RSM is applied to define suitable correlations between the PK parameters and the basal clinical trials.

Once defined, the new equations are substituted in SGM, modifying it. From now on, we will refer to the modified SGM model as MGM (modified Galvanin model).

Following the directions defined in the work by Miao et al. (2011), local sensitivity analysis for model identification is performed, to understand whether the MGM is still locally identifiable or not.

After that, the comparison between the profiles of response produced by the SGM and MGM is carried out to evaluate if the DDAVP can be shortened and redesigned or if it can be completely avoided.

Chapter 3

Preliminary analysis of the model information content

The results of the information content analysis (sensitivity analysis and evaluation of Fisher Information matrix) executed on SGM are presented in this chapter. This analytical procedure is fundamental to determine the starting point for model modification through RSM.

3.1 Results of the sensitivity analysis

The sensitivity analysis has been performed on the SGM to understand the time required by the three kinetic parameters (k_0 , k_1 , k_e) for reaching the maximum in the sensitivity. This analysis allows to visualize the sampling-time range to achieve the highest information that can be gathered from the DDAVP clinical trial. The initial set of parameters has been reported in table 3.1 for each category of disease and healthy subjects considered in the study. The values have been derived considering the average of the kinetic parameters for each category in the pool of subjects. The experimental data for the subjects in the pool are reported in appendix A.

Table 3.1. *Initial set of model parameters.*

	HnonO	HO	2B	Vicenza
k_0	2.87E-02	2.64E-02	1.77E-02	6.66E-02
k_1	2.37E-04	6.25E-04	4.71E-03	1.50E-03
k_e	7.04E-04	1.52E-03	3.23E-03	8.18E-03

The sensitivity analysis has been carried out on three out of five model parameters. Indeed, k_0 , k_1 and k_e are the model kinetic parameters, which aim to be calculated directly from basal clinical trials. The analysis has been executed for both the model responses, antigen VWF:Ag and collagen VWF:CB concentrations, producing a perturbation of 1% on each model parameters one factor at a time.

As is possible to see from figure 3.1, the dynamic sensitivity executed for the VWF:Ag response in HO subjects shows that parameter k_0 can be estimated achieving the maximum of the information around 175 min in the experiment time. The sensitivity of k_1 is set to zero, meaning that the parameter

cannot be estimated from the experimental data. The sensitivity of k_e , instead, does not reach a maximum in the time range of the clinical trial. Therefore, a greater sampling time should be used to get a better estimation of the model parameter.

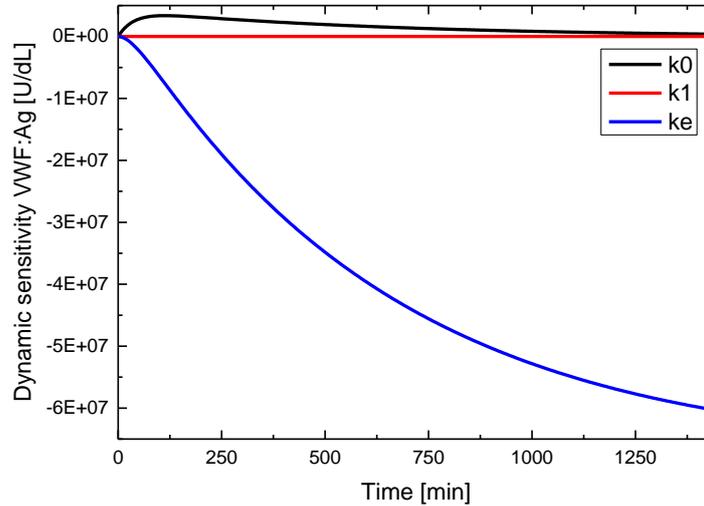


Figure 3.1. Dynamic sensitivity for VWF:Ag response in HO category.

Considering now the dynamic sensitivity on VWF:CB response in figure 3.2, parameter k_0 reaches the maximum in the sensitivity around 175 min as in the previous plot. The proteolytic parameter k_1 can here be identified from the experimental data reaching the maximum in the sensitivity around 500 min, whereas k_e , as before, requires the longest time to achieve its sensitivity peak. This means that the time required for the most informative estimation is the maximum for parameter k_e . Furthermore, neither the sensitivity of k_1 , nor the sensitivity of k_e achieve the stabilization in the experiment time. Hence, a longer sampling time should be advised for the system to stabilize. If the system does not reach stabilization in the experimental time, we are not able to see the entire behaviour of the sensitivity profiles of the parameters and deviations or peaks may not be detectable.

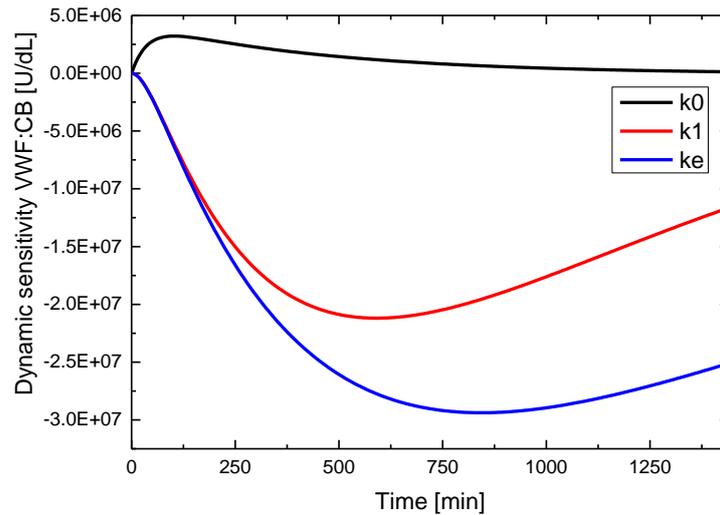


Figure 3.2. *Dynamic sensitivity for VWF:CB response in HO category.*

The same approach has been applied also for HnonO subjects. The dynamic sensitivities of the two model responses are reported in figures 3.3 and 3.4.

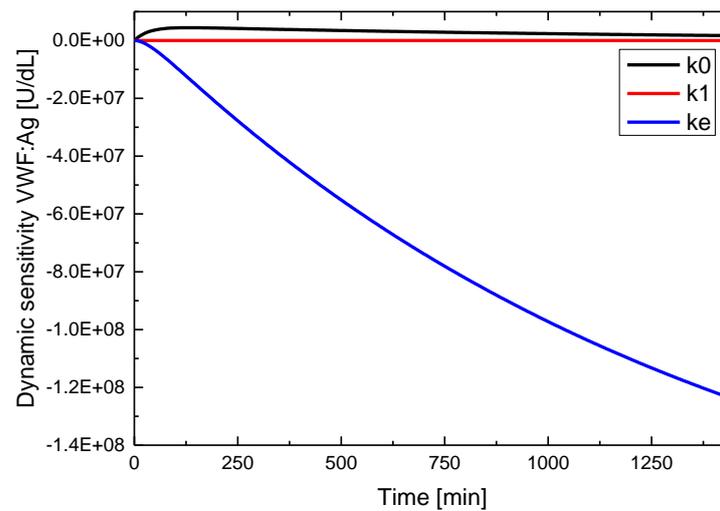


Figure 3.3. *Dynamic sensitivity for VWF:Ag response in HnonO category.*

From figure 3.3 the same conclusions written for VWF:Ag response in the case of HO subjects can be derived. Indeed, parameter k_e requires the longest sampling time and it does not reach the stabilization of the sensitivity at the end of the experimental time. Parameter k_1 instead cannot be identified from the experimental data, whereas parameter k_0 requires the shortest sampling time to achieve the sensitivity peak.

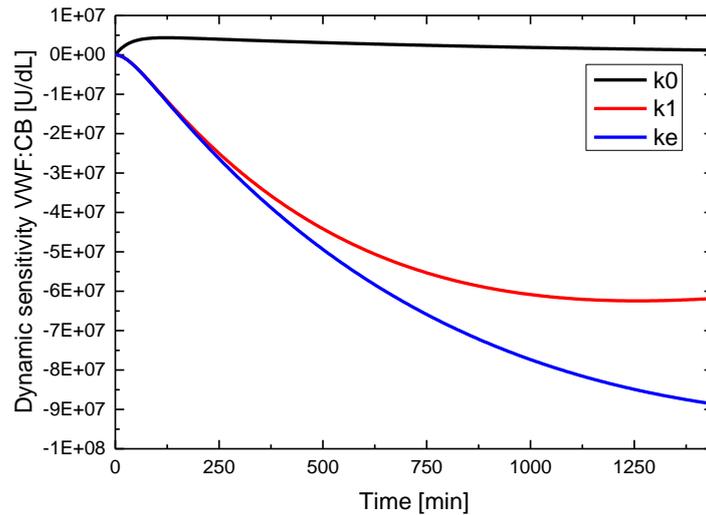


Figure 3.4. Dynamic sensitivity for VWF:CB response in HnonO category.

In case of VWF:CB response in HnonO category, the estimation of the parameters can be performed from the experimental data, but only the dynamic sensitivity of k_0 reaches the peak in the time range of the experiment and the stabilization at the end of the sampling time. Moreover, until 250 min as sampling time, the profiles of parameters k_1 and k_e overlap, meaning that measurements taken here will not allow a distinction between these parameters.

The same conclusions can be derived also analysing the dynamic sensitivity for the unhealthy categories. The results of sensitivity analysis conducted for the antigen and collagen responses in case of 2B category are presented in figures 3.5 and 3.6.

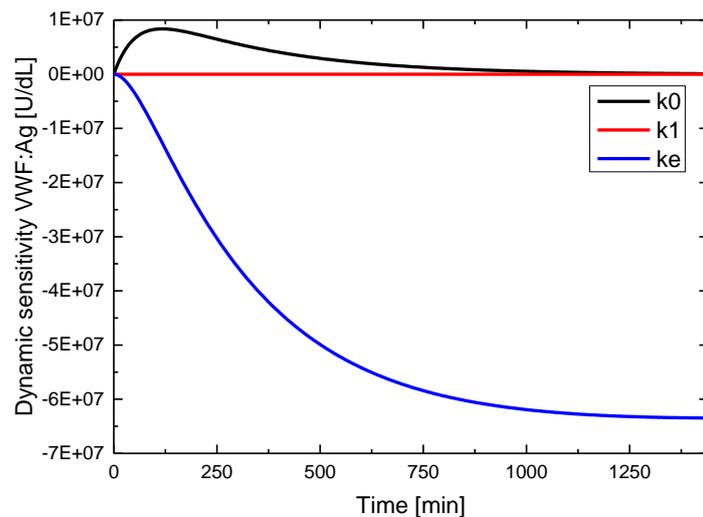


Figure 3.5. Dynamic sensitivity for VWF:Ag response in 2B category.

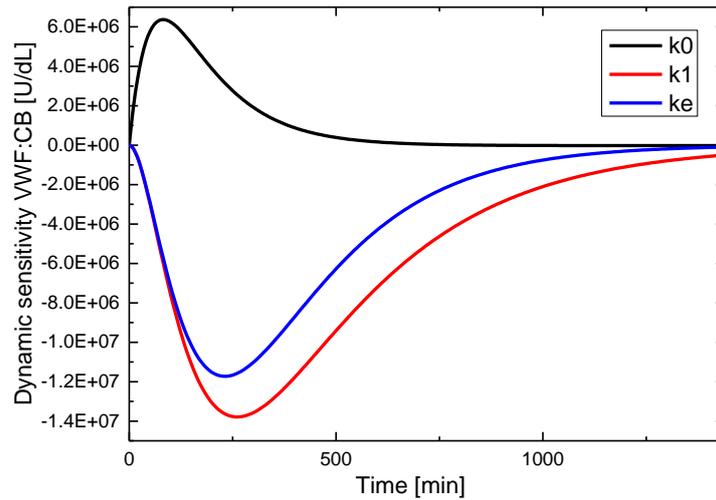


Figure 3.6. Dynamic sensitivity for VWF:CB response in 2B category.

Results show that parameter k_e requires the longest time for achieving the maximum in the sensitivity in case of VWF:Ag response and it does not reach the stabilization of the sensitivity at the end of experimental time. On the other hand, in case of VWF:CB response, all the sensitivities executed for the three parameters achieve the sensitivity peak in the sampling time and the stabilization at the end of the considered time range. Therefore, if samples are taken around the maximum of the profiles parameters k_0 , k_1 and k_e can be uniquely estimated from the experimental data with good precision. However, as is possible to see in the experimental time-range [0 100] of figure 3.6, the sensitivities of parameters k_1 and k_e overlap, meaning that if samples are collected in this range, the two parameters are totally correlated and parameters cannot be estimated.

Again, results of the dynamic sensitivity for Vicenza category are reported in figures 3.7 and 3.8.

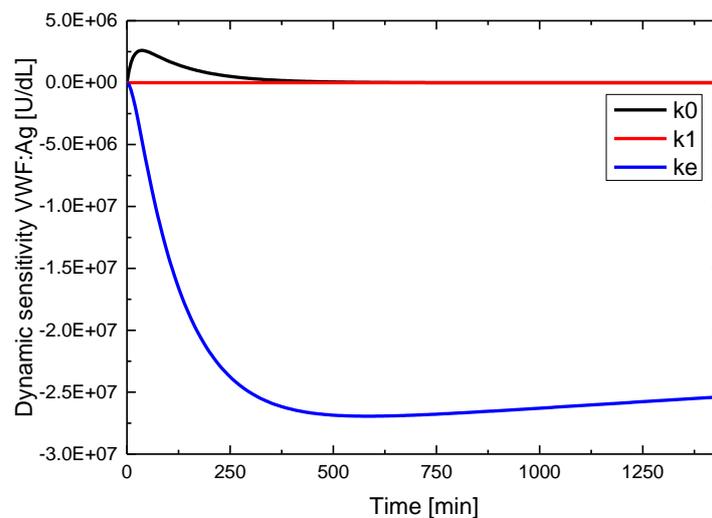


Figure 3.7. Dynamic sensitivity for VWF:Ag response in Vicenza category.

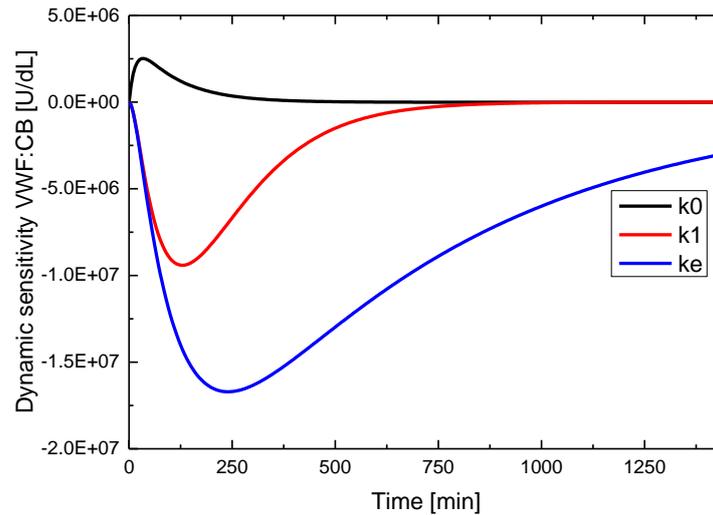


Figure 3.8. *Dynamic sensitivity for VWF:CB response in Vicenza category.*

The same conclusions as for 2B subjects can be derived also analysing the results of Vicenza category. Indeed, in case of VWF:CB response, all the profiles of the sensitivity show the peak in the experimental time. However, as is possible to see from figures 3.7 and 3.8, parameter k_e requires the longest time for reaching the peak in the collagen response, and in both the responses it is not able to get the stabilization at the end of the sampling time.

3.2 Results of FIM (Fisher Information Matrix) evaluation

Analysis of the sensitivity shows where it is possible to have the maximum of the information for each response in each category considering the different model parameters, but, in this way, it is not easy to generalize where it is better to sample in order to maximize the information we can get from the clinical trials. Therefore, more appropriate metric needs to be used for multiple input/multiple output systems like ours. Analysis on the maximum of the information is carried out on the Fisher information matrix (§2.2.1.1). The trace of the Fisher information matrix is one of the most important metric to evaluate the region where to sample in order to get the maximum in the information for improving parameter estimation. In figures 3.9, 3.10, 3.11 and 3.12 the trace of FIM on the two responses for each category of disease and healthy subjects is reported.

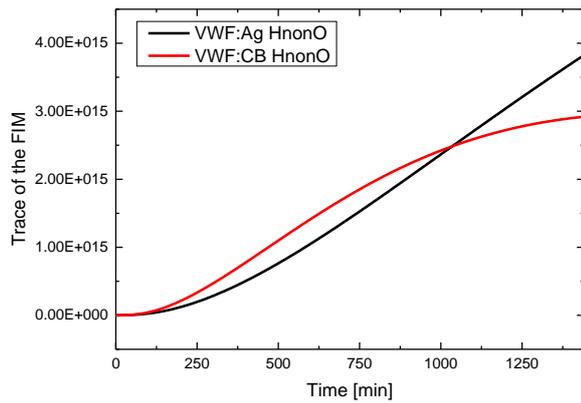


Figure 3.9. Trace of FIM in HnonO category.

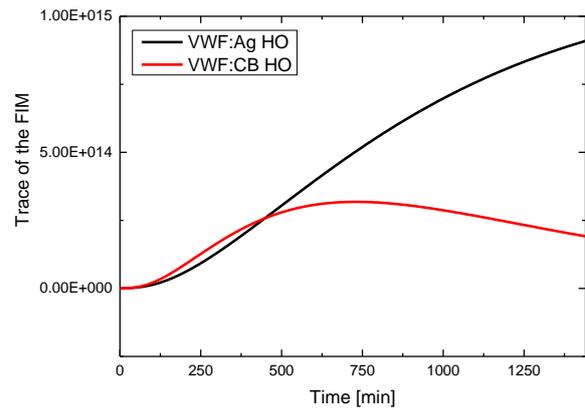


Figure 3.10. Trace of FIM in HO category.

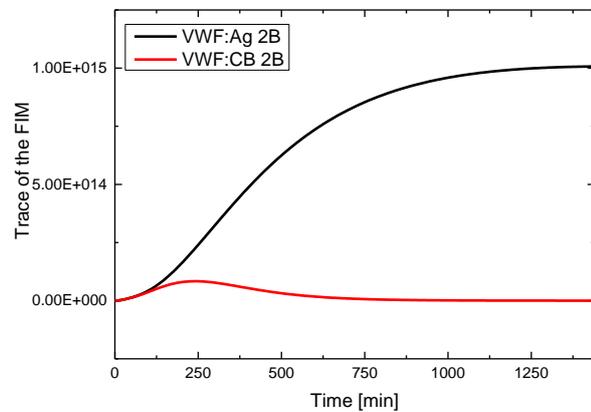


Figure 3.11. Trace of FIM in 2B category.

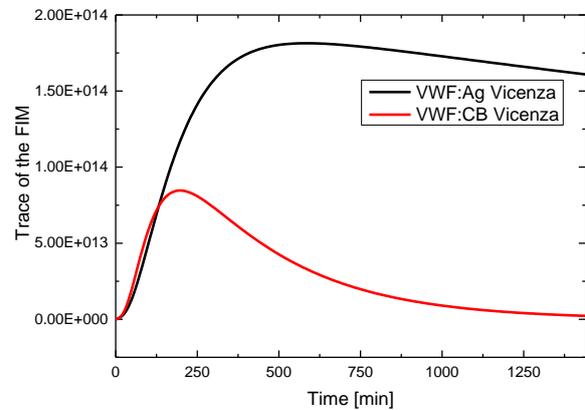


Figure 3.12. Trace of FIM in Vicenza category.

As is possible to see from the figures, the trace does not reach the maximum in the information content for all the categories with the exception of Vicenza category considering the antigen concentration. This behaviour has been produced by the trend of the sensitivity on parameter k_e . On the other hand, the peak in the information content is achieved considering the collagen response in all the cases except for HnonO category, for which a longer experiment time should be advised.

As is possible to understand from this preliminary analysis k_1 and k_e requires the longest time for achieving the sensitivity peak in the DDAVP execution. Moreover, the elimination constant k_e does not even reach the maximum of the information content in the 24 hours of DDAVP execution. Therefore, to achieve the goal of the project, that is to reduce the DDAVP execution time, it seems reasonable to work on the definition of suitable correlations able to calculate k_1 and k_e parameters using only basal values derived from standard clinical trials. Furthermore, from the results obtained, the release parameter k_0 requires the minimum time for reaching the sensitivity peak. However, to tackle the ambitious target of the research, that is to avoid the DDAVP execution, a basal state

correlation should be defined also for the release parameter k_0 . The procedure applied in the definition of the suitable basal state relations for the calculation of the kinetic parameters are reported in the following chapter.

Chapter 4

Model modification and validation

4.1 Results of model modification

In this chapter, RSM is applied to find basal state correlations between k_0 , k_1 , k_e parameters and standard clinical trials (§1.3.1). Once defined, the new equations are substituted in the SGM developing the MGM, with the target of reducing the time or avoid the DDAVP execution. The general procedure that has been followed for the development of the response surfaces is described in figure 4.1.

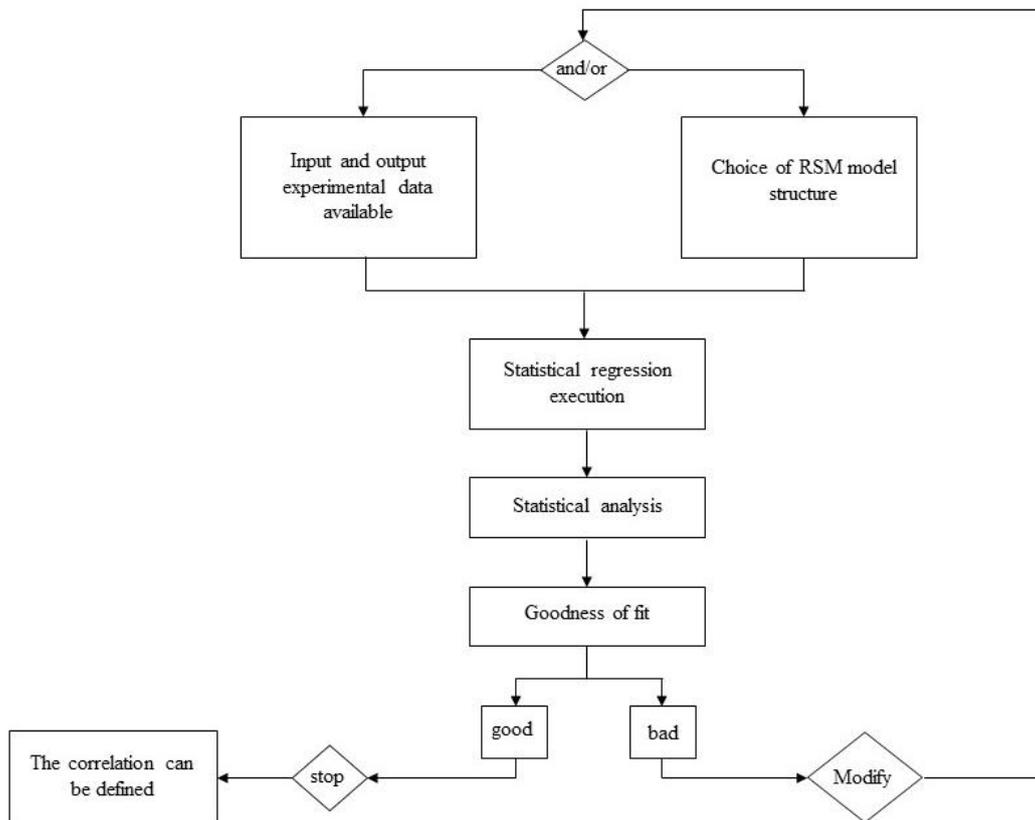


Figure 4.1. RSM procedure.

4.1.1 RSM for k_e

The information content analysis (§3) confirmed clearly that the SGM is uniquely identifiable and the PK parameters might be estimated from the experimental data. However, the sensitivity analysis demonstrated that the time required to achieve the maximum in the information for parameter k_e is the highest compared to the other kinetic parameters (k_0, k_1). Therefore, it seems clearly interesting to investigate an explicit correlation, which allows to calculate the elimination constant k_e from basal clinical trials, without considering the DDAVP test. The correlation has been investigated through the RSM, developing a so called black-box model, which considers k_e as output variable and the two physiological ratios VWFpp ratio and VWF:R (derived by the three basal clinical trials VWF:Ag, VWF:CB, VWFpp) as inputs. The basal values of the three clinical trials for each subject in each category of disease considered in the research work have been supplied by the Hospital of Padova; whereas the values of k_e for each patient have been estimated with the procedure described in §2.1.1. Patients' values of the above defined quantities, required for the development of the regression model, are reported in appendix A. The collaboration with the medical school has been fundamental in the definition of the response surfaces. Indeed, the physiology of VWF represents a good starting point for finding reliable correlations. Preliminary studies and data analysis have been conducted to define the best regression model, whose average AIC is 40.3 ± 5.36 . The mathematical form of the most performant surface of fitting is the linear response surface with interactions (eq. 4.1):

$$k_e = A + B \cdot \text{VWFpp ratio} + C \cdot \text{VWF:R} + D \cdot \text{VWFpp ratio} \cdot \text{VWF:R} \quad (4.1)$$

The determination of the right correlation is the result of a long set of experiments, which have been carried out to find the most significant relations between the model parameters and the basal clinical trials. The overall procedure is reported in appendix B (§B.1).

The fittings of the experimental data with the linear response surface with interactions are illustrated in figures 4.2, 4.3, 4.4 and 4.5, for the different categories.

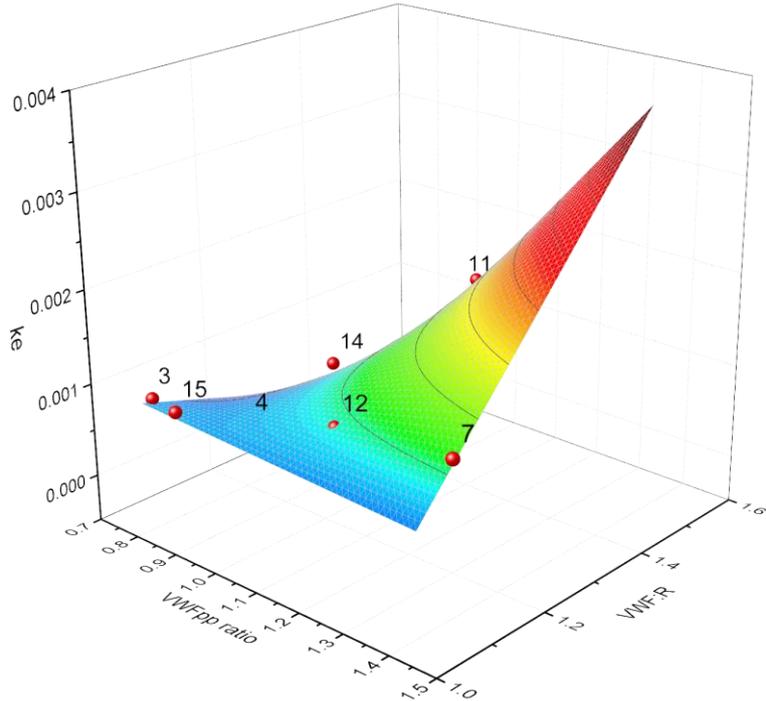


Figure 4.2. Linear response model surface with interactions considering HnonO subjects (subjects removed: 1, 9, 13, 10, 8, 6).

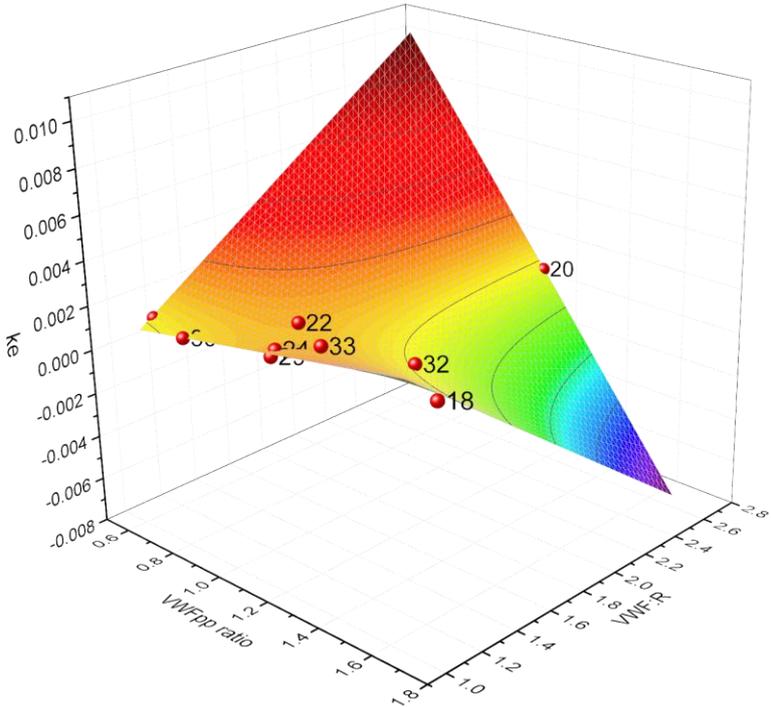


Figure 4.3. Linear response model surface with interactions considering HO subjects (subjects removed: 17, 19, 25, 26, 28, 31).

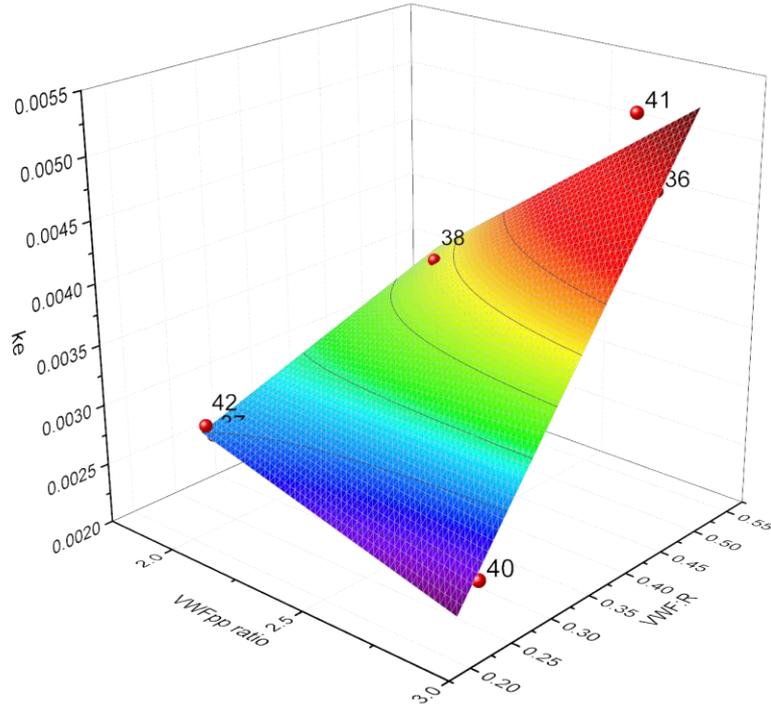


Figure 4.4. Linear response model surface with interactions considering 2B subjects (subject removed: 39).

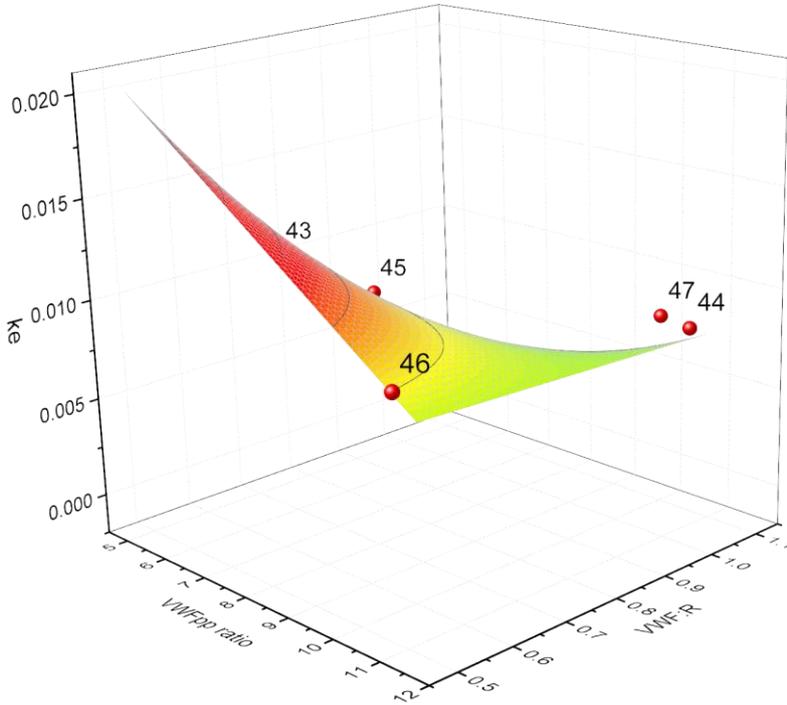


Figure 4.5. Linear response model surface with interactions considering Vicenza subjects (subjects removed: 48, 49).

As is visible in the response surface images, some subjects were removed from the pool to improve the result of the fitting, because they behaved as outliers. The abnormal space position of those subjects is associated to medical factors that arose during the DDAVP execution causing irreversible deviations from the class of belonging. Precisely, some of the removed subjects fainted or collapsed during the DDAVP execution or they are affected by additional blood-coagulation disorders that produce an alteration on the VWF levels. Moreover, some of the abnormal subjects, belonging to 2B or Vicenza categories, possess other mutations on the VWF gene in addition to those expected, therefore a deviation occurs.

The relative error of the surface fitting in the four considered categories is illustrated in figures 4.6, 4.7, 4.8 and 4.9. The quality of the fitting can be analysed through the statistics presented in tables 4.1, 4.2, 4.3 and 4.4 for HnonO, HO, 2B and Vicenza subjects, respectively.

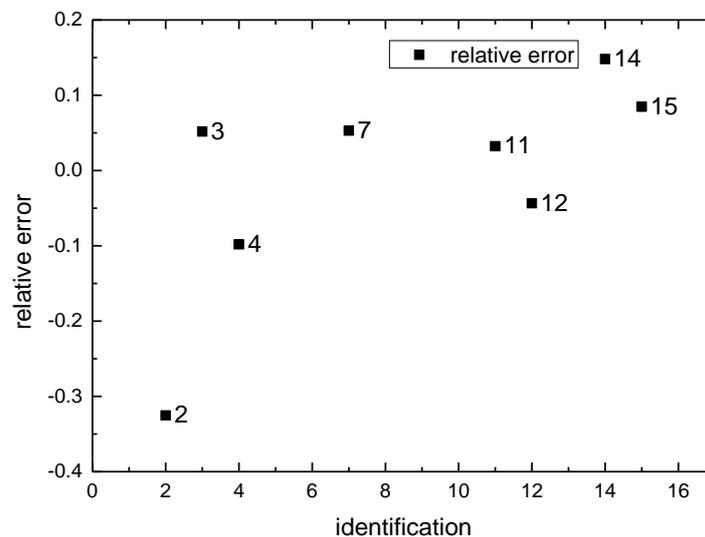


Figure 4.6. Relative error in linear response model fitting with interactions for HnonO subjects (subjects removed: 1, 9, 13, 10, 8, 6).

Table 4.1. Statistics of the fitting with HnonO subjects (subjects removed: 1, 9, 13, 10, 8, 6).

HnonO statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	0.01468	0.00498	2.94712	0.04209	8.50E-04	2.85E-02
B	-0.01521	0.00508	-2.99186	0.04026	-2.93E-02	-0.00109
C	-0.01316	0.0045	-2.92471	0.04304	-2.57E-02	-6.67E-04
D	0.01438	0.00458	3.13695	0.03495	0.00165	2.71E-02
Number of points	8					
Degrees of Freedom	4					
Reduces Chi-Squared	2.02E-08					
Residual Sum of Squares	8.09E-08					
R Value	0.9257					
Adj. R-Square	0.74962					

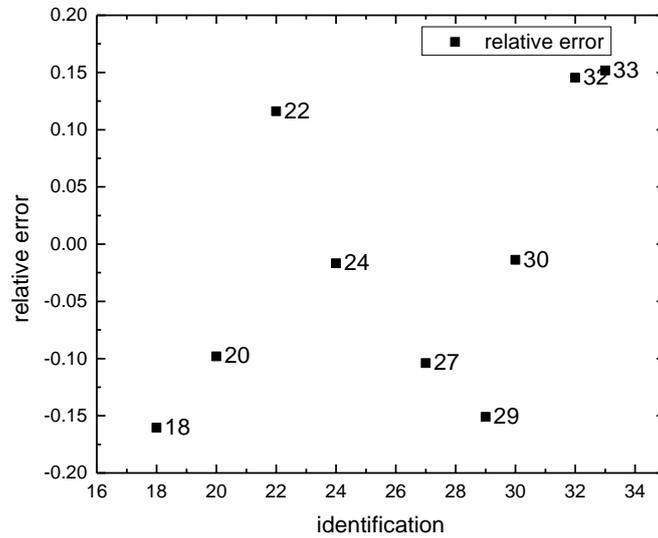


Figure 4.7. Relative error in linear response model fitting with interactions for HO subjects (subjects removed: 17, 19, 25, 26, 28, 31).

Table 4.2. Statistics of the fitting with HO subjects (subjects removed: 17, 19, 25, 26, 28, 31).

HO statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	-0.01397	0.00503	-2.77471	0.03915	-0.02691	-0.00103
B	0.01484	0.00448	3.31547	0.02111	0.00333	0.02634
C	1.31E-02	4.36E-03	2.99477	0.03028	1.85E-03	2.43E-02
D	-0.01209	0.00382	-3.16506	0.02495	-0.02192	-0.00227
Number of points	9					
Degrees of Freedom	5					
Reduces Chi-Squared	9.25E-08					
Residual Sum of Squares	4.62E-07					
R Value	0.90983					
Adj. R-Square	0.72446					

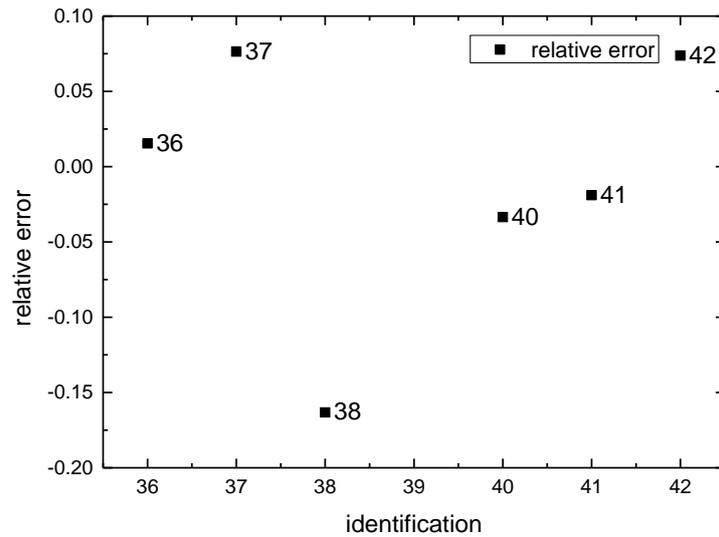


Figure 4.8. Relative error in linear response model fitting with interactions for 2B subjects (subject removed: 39).

Table 4.3. Statistics of the fitting with 2B subjects (subject removed: 39).

2B statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	0.0093	0.00286	3.25126	0.08299	-0.00301	0.0216
B	-0.00348	0.00122	-2.85955	0.10363	0.00873	0.00176
C	-0.02435	0.00846	-2.87773	0.10252	-0.06075	0.01206
D	0.01299	0.00354	3.6745	0.06673	-0.00222	0.02821
Number of points	6					
Degrees of Freedom	2					
Reduces Chi-Squared	1.35E-07					
Residual Sum of Squares	2.69E-07					
R Value	0.97905					
Adj. R-Square	0.89634					

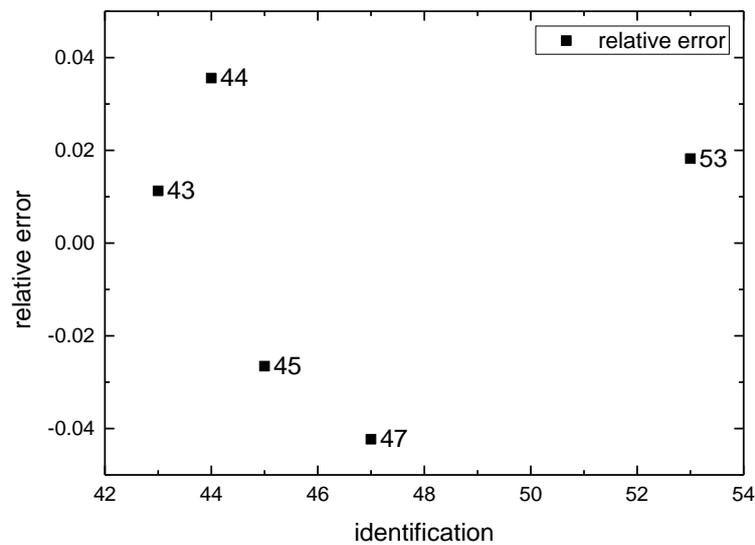


Figure 4.9. Relative error in linear response model fitting with interactions for Vicenza subjects (subjects removed: 48, 49).

Table 4.4. *Statistics of the fitting with Vicenza subjects (subjects removed: 48, 49).*

Vicenza statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	0.02005	0.01106	1.8133	0.32084	-0.12045	0.16055
B	-0.00162	0.00181	-0.89051	0.53683	-0.02466	0.02143
C	-0.00916	0.01158	-0.79152	0.57375	-0.15626	0.13794
D	0.00138	0.00177	0.77778	0.57917	-0.02112	0.02387
Number of points	5					
Degrees of Freedom	1					
Reduces Chi-Squared	3.19E-07					
Residual Sum of Squares	3.19E-07					
R Value	0.92946					
Adj. R-Square	0.45561					

The statistics are generated by the *Nonlinear surface fit analyser package* in OriginPro software. Table 6 has been reported to summarize the most important statistics produced by the analysis.

Table 4.5. *Summary of goodness of fit statistics.*

	DoF	Reduced chi-sqr	Residuals sum of squares	R^2	\bar{R}^2
HnonO	4	2.02E-08	8.09E-08	0.9257	0.74962
HO	5	9.25E-08	4.62E-07	0.90983	0.72446
2B	2	1.35E-07	2.69E-07	0.97903	0.89634
Vicenza	1	3.19E-07	3.19E-07	0.92946	0.45561

As is possible to understand from the results of the statistics reported in table 4.5 and from the fitting illustrated in the images, the regression model successfully interpolates the experimental data. The R^2 is indeed higher than 90% in all the considered categories indicating a good quality of the fitting. Moreover, as is clearly shown in tables 4.1, 4.2, 4.3 and 4.4, the t-value is higher than the chi-squared of reference stating that the experimental points are sufficient for the development of a reliable response surface; also, the confidence intervals show meaningful values in line with a high quality of the fitting.

The new explicit correlation for parameter k_e , which has been found statistically reliable, produces a deep reduction on the DDAVP execution time and this statement finds clear confirmation in the following section.

4.1.1.1 SGM modification and simulation of MGM_1

The SGM has been modified (figure 4.10) adding in the equations set the new explicit correlation for k_e (eq. 4.1). This modification leads to the development of MGM_1 (modified Galvanin model, first version). The algebraic relation introduces four new model parameters (A, B, C, D), whose values have been obtained from the fitting procedure executed in OriginPro[®]. The values of the parameters for each category considered in the study are summarized in table 4.6.

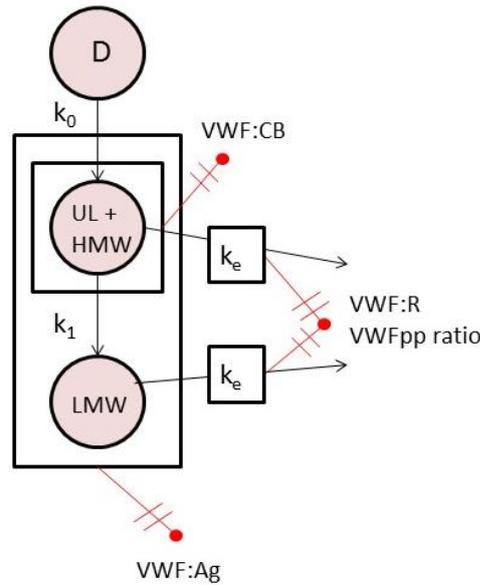


Figure 4.10. MGM_1 scheme.

Table 4.6. Model parameters of the RSM.

	HnonO	HO	2B	Vicenza
A	0.01468	-1.20E-02	0.0093	0.02005
B	-0.01521	1.31E-02	-0.00348	-0.00162
C	-0.01316	0.01089	-0.02435	0.00916
D	0.01438	-1.03E-02	0.01299	0.00138

The MGM_1 can be represented as in figure 4.10, in which it is shown how the value of k_e is determined by two basal quantities (VWF:R and VWFpp ratio), whereas k_0 and k_1 still need to be estimated by the DDAVP. To compare the profiles generated by the SGM (§2.1) and MGM_1, the average subject for each category has been taken as reference. The average input data for each category, required to simulate MGM_1, are reported in table 4.7.

Table 4.7. *VWFpp ratio and VWF: R average values in the considered categories.*

	VWFpp ratio	VWF: R
HnonO	1.040	1.252711
HO	1.113	1.258576
2B	2.320	0.363891
Vicenza	0.881	8.843210

The comparison between the average k_e values calculated with the new explicit correlation and the average k_e estimated values in the SGM are presented in table 4.8.

Table 4.8. *Calculated average k_e values and the average estimated k_e values.*

	k_e calculated by MGM_1	k_e estimated through SGM	relative error
HnonO	6.73E-04	7.04E-04	0.046062407
HO	0.001858	0.00152	0.181916039
2B	0.003332	0.00323	0.030612245
Vicenza	0.00840673	0.00818	0.026970059

As is possible to read from table 4.8, the calculated and the estimated values of k_e are numerically close within each other, indeed the relative error is lower than 10% in all the categories except for HO category. This can be due to the intrinsic internal variability of the pool of subjects considered in the study for the HO category. This finding reflects on the profiles of VWF:Ag and VWF:CB responses. From figures 4.11, 4.12, 4.13 and 4.14, it is clearly evident that the profiles of SGM and MGM_1 overlap in each category with exception of HO category, but the deviation ($< \pm 20$ U/dL) produced in the value of the peak of the two model responses VWF:Ag and VWF:CB can be accepted by the medical community.

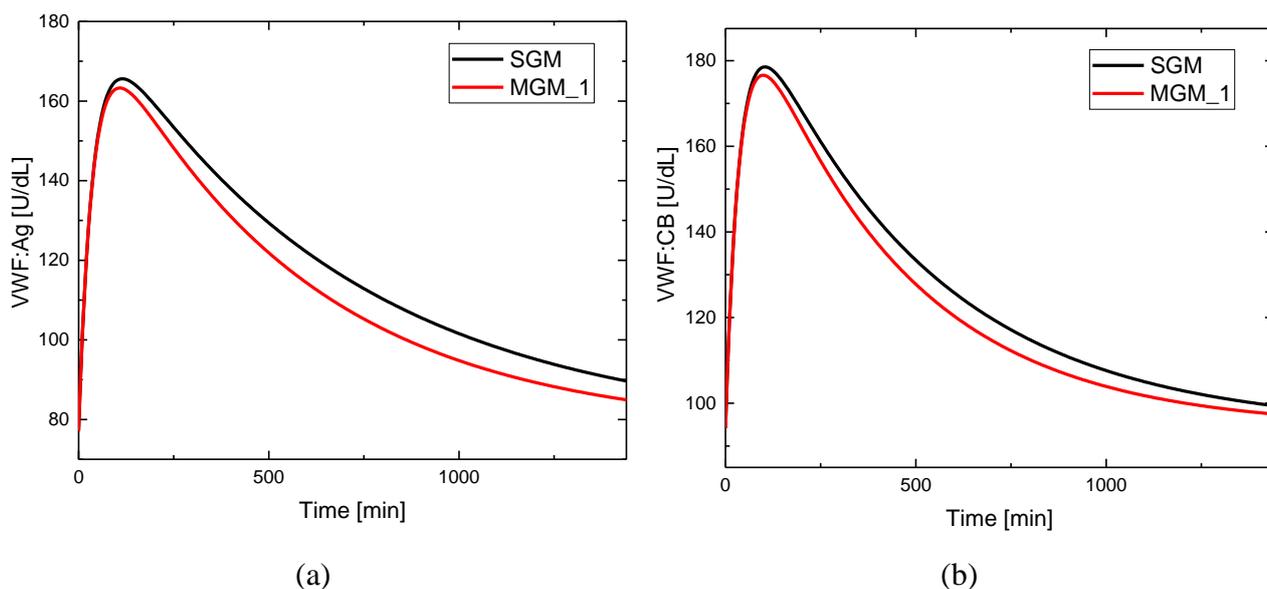


Figure 4.11. Simulated response VWF:Ag (a) and VWF:CB (b) with MGM_1 and SGM in HO category.

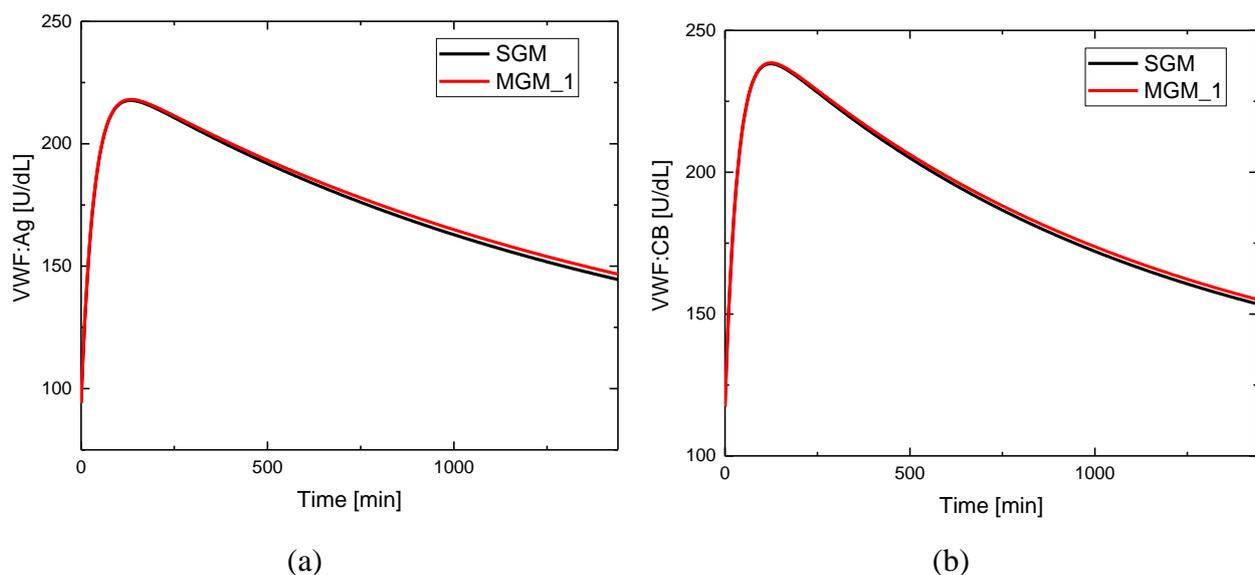


Figure 4.12. Simulated response VWF:Ag (a) and VWF:CB (b) with MGM_1 and SGM in HnonO category.

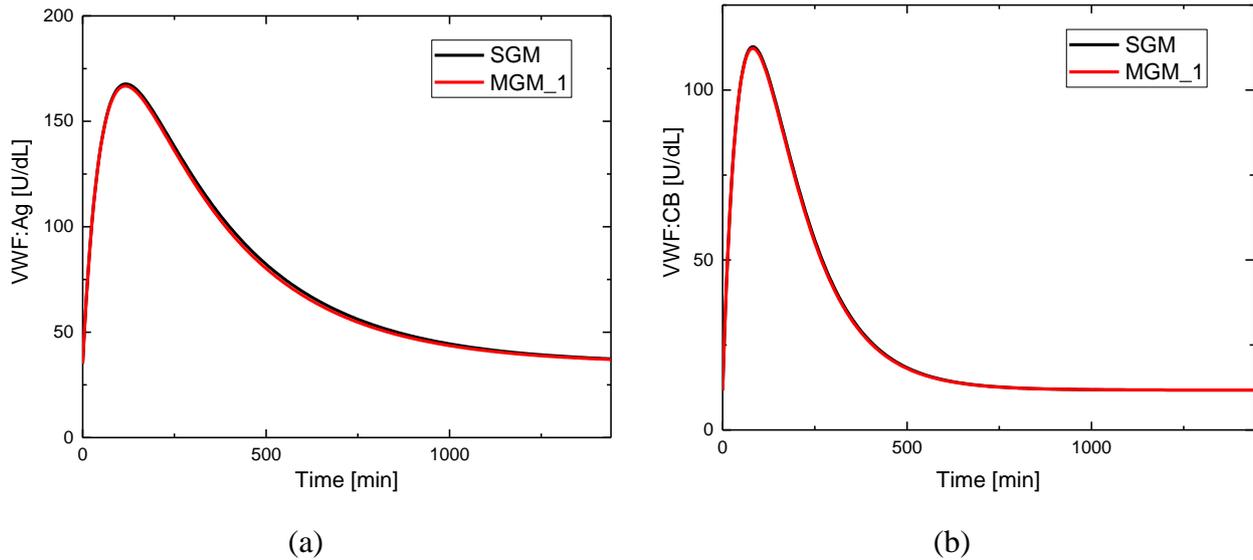


Figure 4.13. Simulated response VWF:Ag (a) and VWF:CB (b) with MGM_1 and SGM in 2B category.

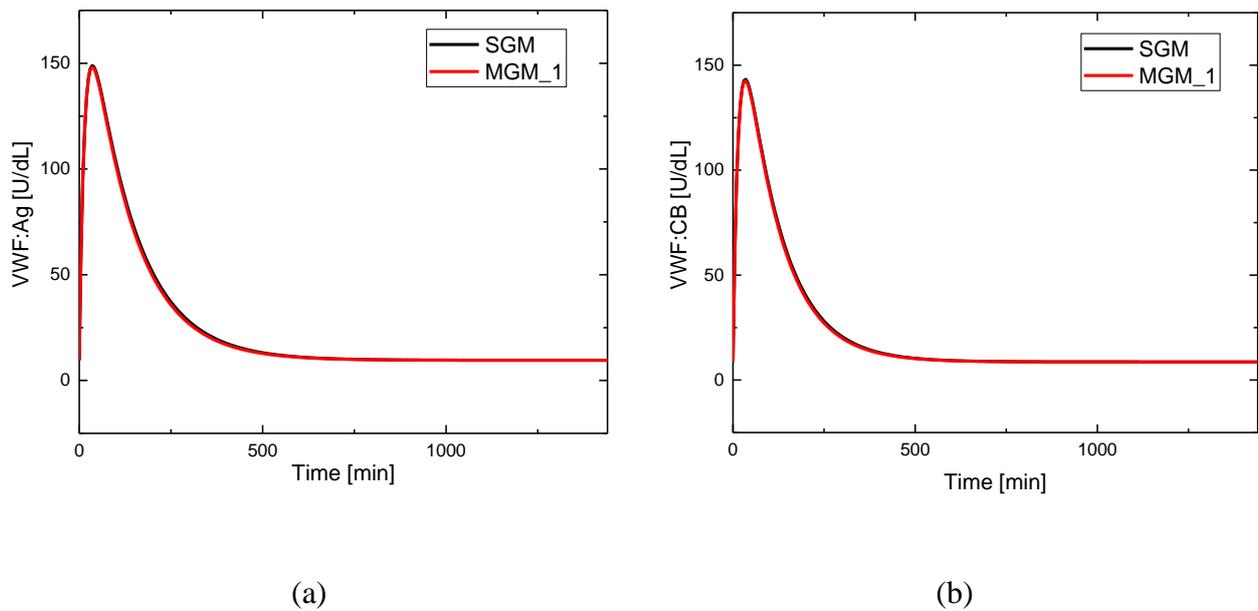


Figure 4.14. Simulated response VWF:Ag (a) and VWF:CB (b) with MGM_1 and SGM in Vicenza category.

4.1.1.2 Information content analysis on MGM_1

Information content analysis has been executed in order to evaluate whether MGM_1 is still locally identifiable.

The local sensitivity analysis has been conducted in gPROMS[®] acting a perturbation of 1% on the model parameters, which mainly characterize MGM_1.

In particular, the vector of the modified model parameters can be represented as a two dimensional array:

$$\theta = [k_0, k_1]$$

The sensitivity analysis has been conducted for each parameter in the vector and for each category of disease or healthy subjects. Results of the two model responses VWF:Ag and VWF:CB are reported in figures 4.15 and 4.16 for the healthy subjects HnonO and HO, whereas the profiles of the sensitivities in the 2B and Vicenza categories are presented in figures 4.17 and 4.18, respectively.

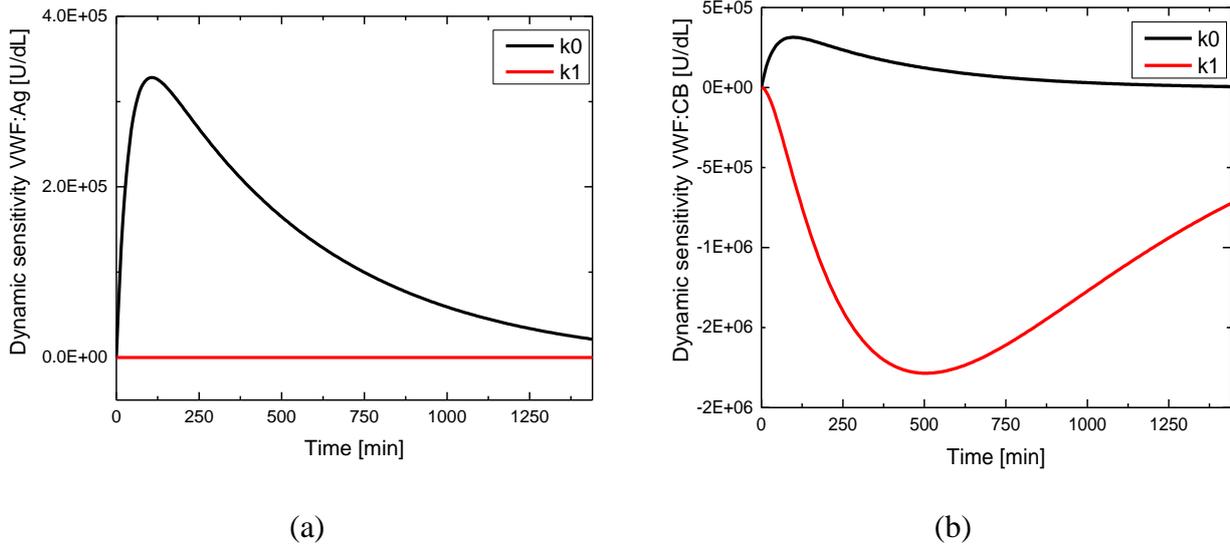


Figure 4.15. Dynamic sensitivity for VWF:Ag (a) and VWF:CB (b) responses in HnonO category for parameters k_0 and k_1 .

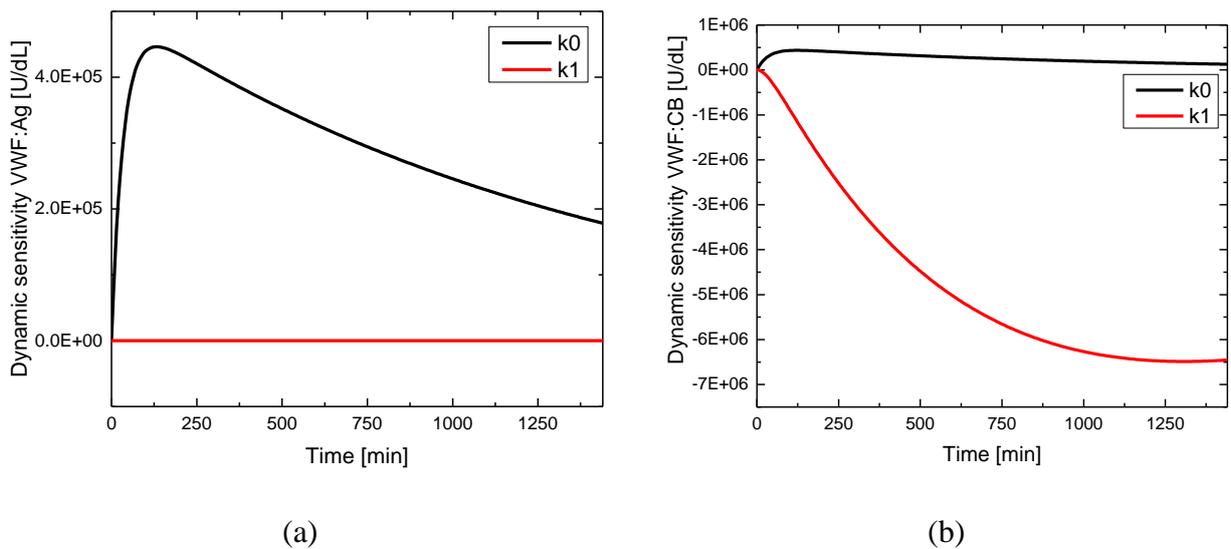


Figure 4.16. Dynamic sensitivity for VWF:Ag (a) and VWF:CB (b) responses in HO category for parameters k_0 and k_1 .

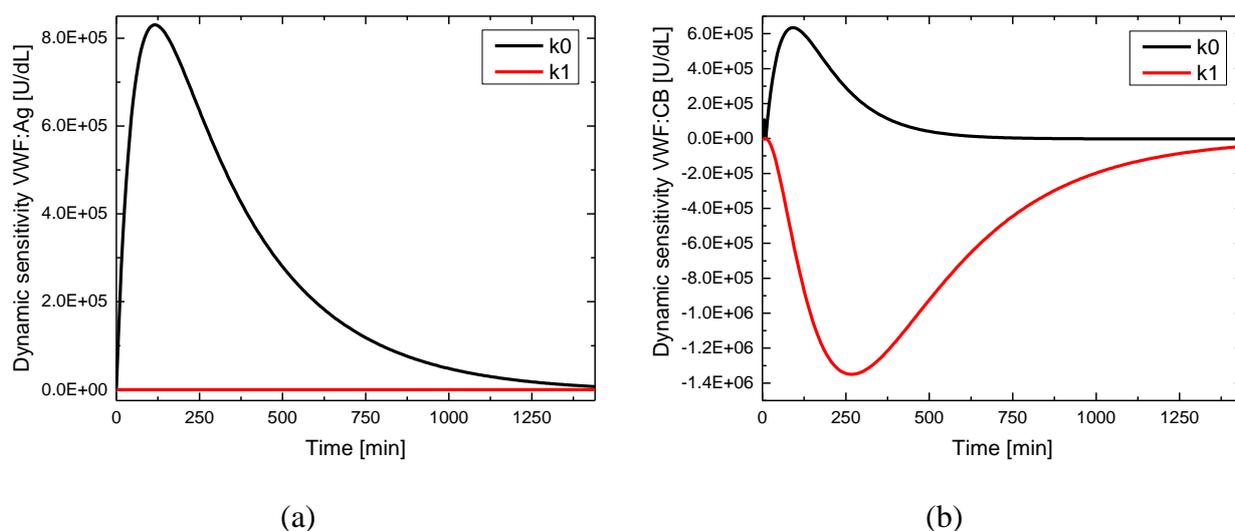


Figure 4.17. Dynamic sensitivity for VWF:Ag (a) and VWF:CB (b) responses in 2B category for parameters k_0 and k_1 .

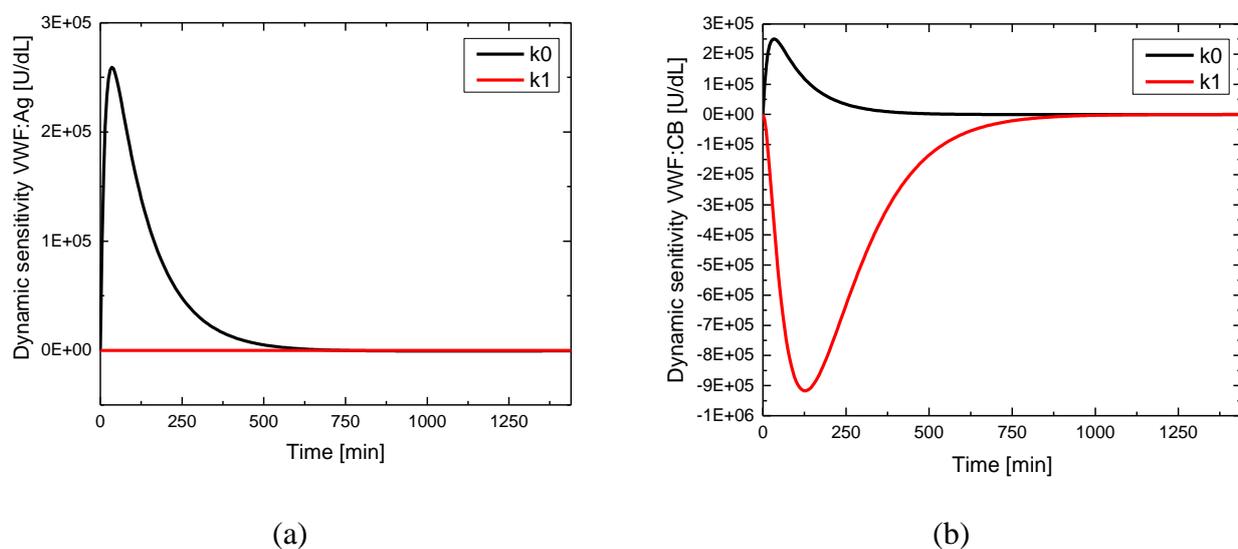


Figure 4.18. Dynamic sensitivity for VWF:Ag (a) and VWF:CB (b) responses in Vicenza category for parameters k_0 and k_1 .

It is important to consider that also the sensitivity for VWFpp ratio has been executed in the modified model being a countercheck for the time required by the identification of the elimination constant k_e in SGM.

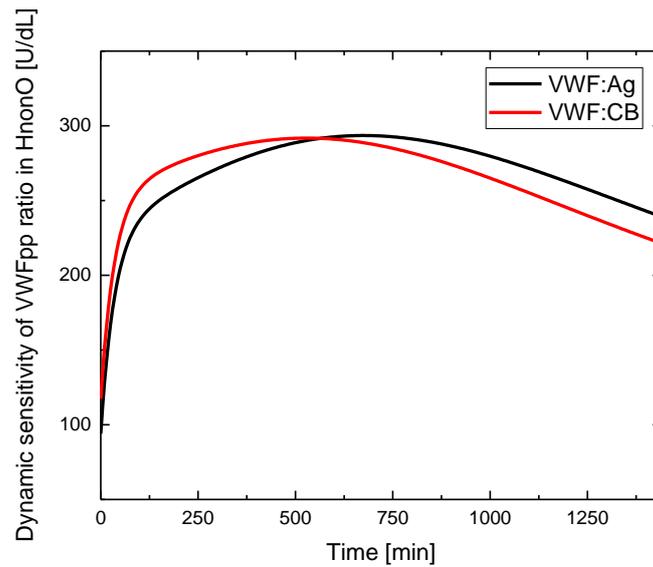


Figure 4.19. Dynamic sensitivity for VWF:Ag and VWF:CB responses in HnonO category on VWFpp ratio.

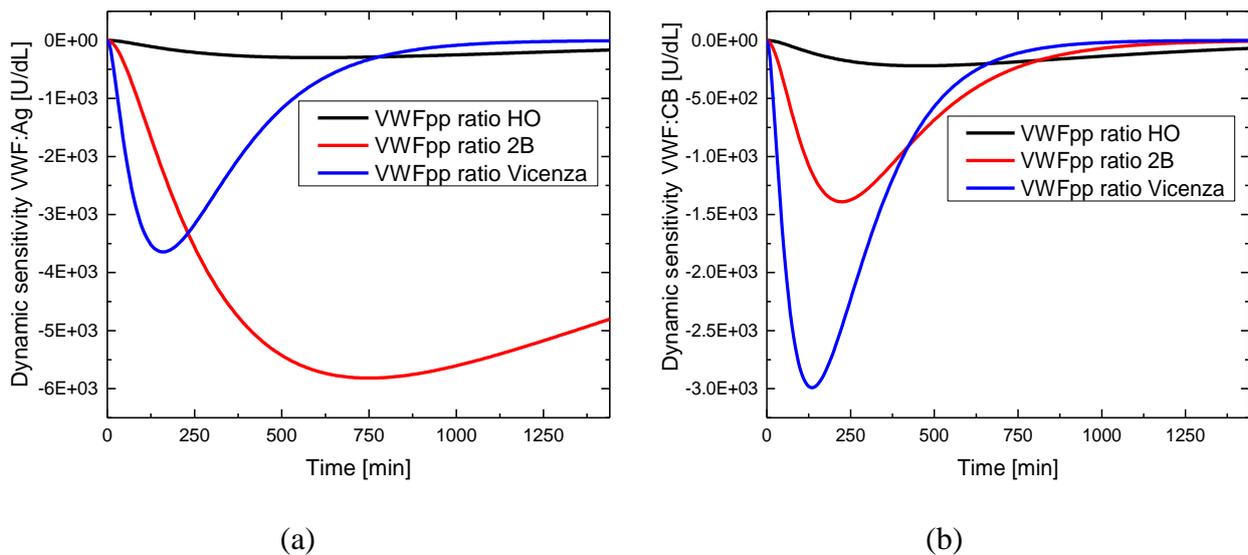


Figure 4.20. Dynamic sensitivity for VWF:Ag (a) and VWF:CB (b) responses in HO, 2B and Vicenza categories on VWFpp ratio.

As it is possible to understand from the profiles in each category (figures 4.15, 4.16, 4.17, 4.18), the parameter k_1 cannot be identified by VWF:Ag experimental data, whereas it can be identified by VWF:CB response. Parameter k_0 instead peaks in both the model responses, but the highest value of the peak is associated to the antigen concentration in all the categories, meaning that VWF:Ag response is more informative for the estimation of the release kinetic parameter.

Hence, DDAVP clinical trial needs to be carried out for the estimation of the two model parameters k_0 and k_1 . However, as the plots illustrate, the execution time required to achieve the sensitivity peak

of the two parameters (k_0 , k_1) in MGM_1 is shorter (almost halved) compared to the identification time needed by the elimination constant k_e in SGM. To stress out the improvement, the sensitivities for each category on VWFpp ratio have been reported in figures 4.19 and 4.20 (two figures are required to represent the profiles between the different categories because scales are different). VWFpp ratio is clinically used to indirectly quantify the elimination constant k_e . The profiles of the sensitivity on VWFpp ratio produced by MGM_1 show almost the same trends of the sensitivity on k_e in SGM. Figures 4.19 and 4.20 have been inserted in the discussion to confirm that MGM_1 is clearly able to represent the physiology of the system, evaluating the elimination constant k_e with VWFpp ratio quantity, in agreement with medical literature. However, we are not interested in the time required by VWFpp ratio to get the highest sensitivity value, because only its basal value, together with that of VWF:R, is used for the calculation of the elimination constant k_e through the new explicit correlation (eq. 4.1).

4.1.2 RSM for k_1

As demonstrated in §4.1.2, the definition of a correlation for parameter k_e theoretically allows to halve the DDAVP execution time. However, still 8-12 hours are required to achieve a statistical satisfactory estimation of the proteolytic parameter k_1 through the DDAVP clinical trial. Therefore, to improve the reduction of the DDAVP execution time a suitable correlation for parameter k_1 has been investigated. Several basal state clinical quantities (i.e. VWFpp ratio, VWF:R and VWF:Rco) can be related to the proteolytic parameter. As reported in appendix B.2, the investigation of the right correlation for parameter k_1 has been challenging. Indeed, not only the quality of the fitting and the physiological meaning of the equation need to be considered, but also the clinical trials that have to be executed must be easy to conduct. This is fundamental to respect the concept of simplification that is at the core of the research goal.

In collaboration with the medical school, we deduced that the proteolytic parameter k_1 is related to VWF:Ag, which represents the number of antigens in the blood stream, and to the elimination that can be measured by the VWFpp ratio at the basal state. Therefore, reasonably k_1 can be expressed as function of VWF:Ag and VWFpp ratio. The linear response surface with interactions for the proteolytic parameter is mathematically defined as follow:

$$k_1 = A + B \cdot \text{VWF: Ag} + C \cdot \text{VWFpp ratio} + D \cdot \text{VWF: Ag} \cdot \text{VWFpp ratio} \quad (4.2)$$

The correlation (eq. 4.2) seems to work well as results demonstrate. The fitting with the linear response surfaces with interactions are reported in figures 4.21, 4.22, 4.23 and 4.24 for HnonO, HO, 2B and Vicenza categories, respectively.

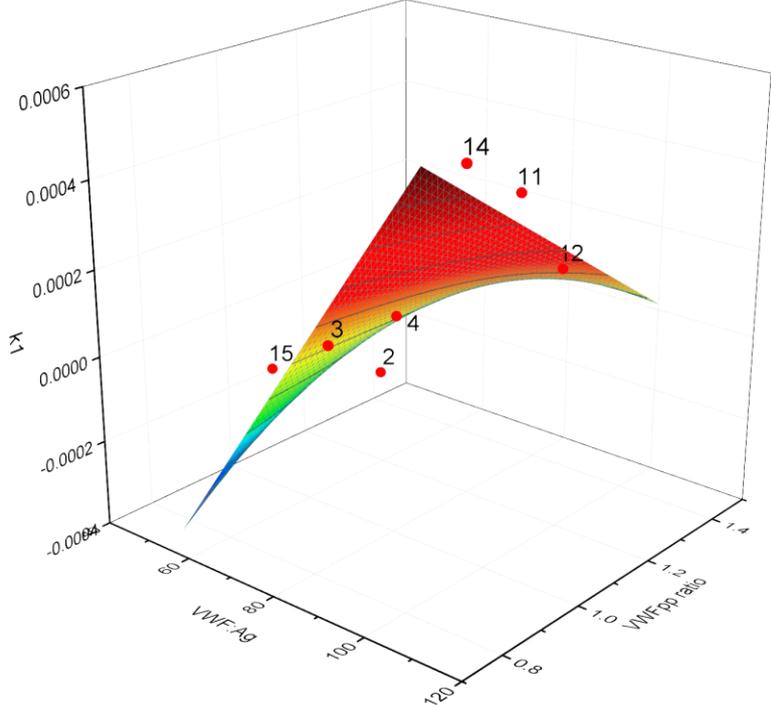


Figure 4.21 Linear response model surface with interactions considering HnonO subjects (subjects removed: 1, 9, 13, 10, 8, 6).

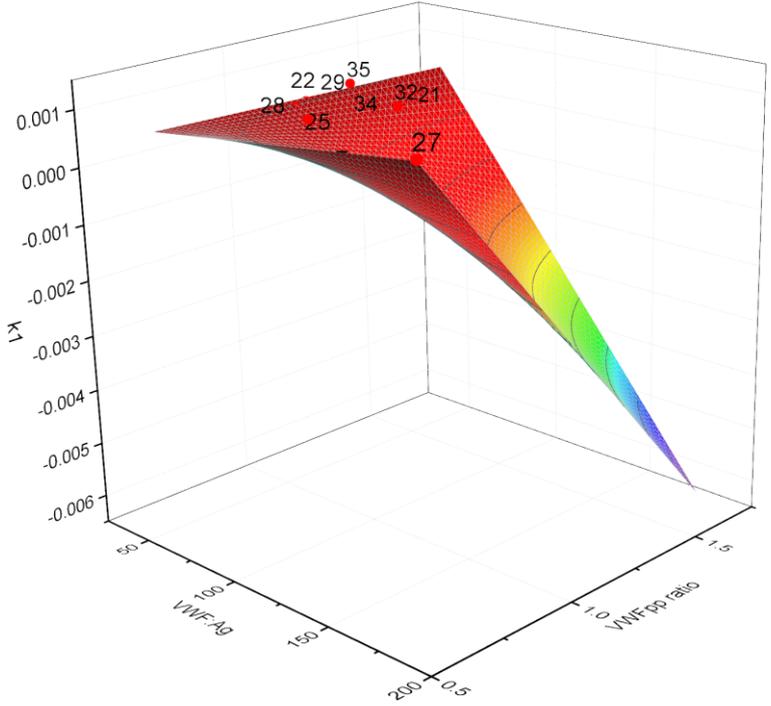


Figure 4.22. Linear response model surface with interactions considering HO subjects (subjects removed: 17, 19, 25, 26, 28, 31).

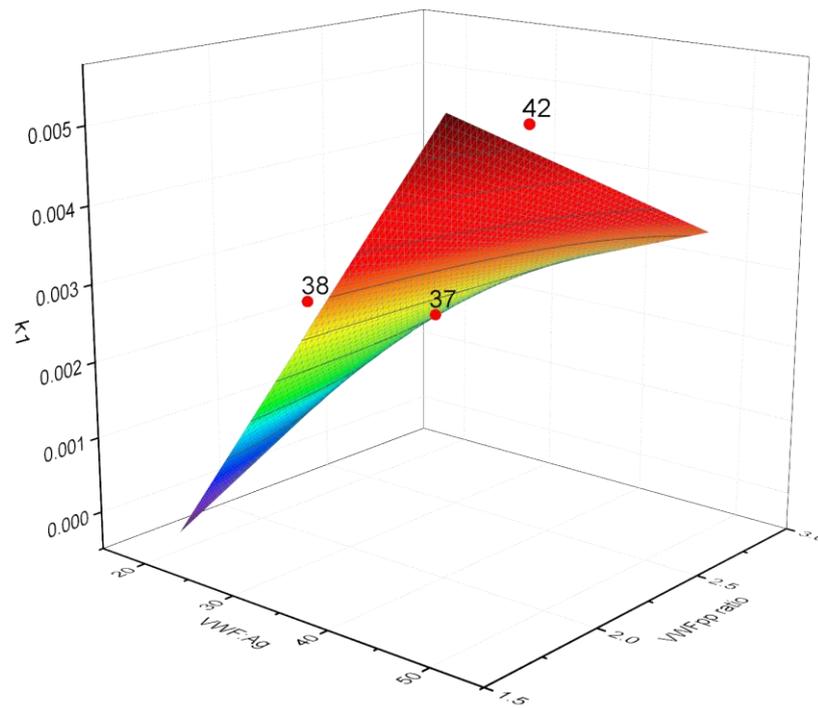


Figure 4.23 Linear response model surface with interactions considering 2B subjects (subject removed: 39).

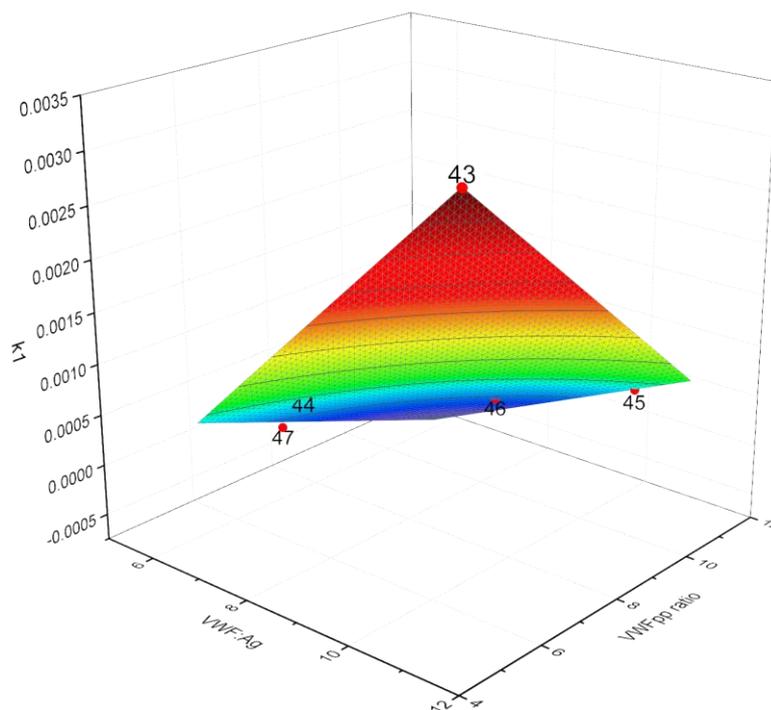


Figure 4.24 Linear response model surface with interactions considering Vicenza subjects (subjects removed: 48, 49).

The quality of the fitting is confirmed also by the statistical results reported in tables 4.9, 4.10, 4.11 and 4.12 for HnonO, HO, 2B and Vicenza category, respectively.

Table 4.9. *Statistics of the fitting with HnonO subjects (subjects removed: 1, 9, 13, 10, 8, 6).*

HnonO statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	-0.00312	0.00103	3.039	0.038	0.006	-2.71E-04
B	4.09E-06	1.36E-05	3.002	0.040	3.08E-06	7.87E-05
C	0.02233	8.82E-04	2.636	0.057	-1.23E-04	0.005
D	-2.96E-05	1.23E-05	-2.408	0.073	-6.37E-05	4.25E-06
Number of points	8					
Degrees of Freedom	4					
Reduces Chi-Squared	1.11E-08					
Residual Sum of Squares	4.43E-08					
R Value	0.92					
Adj. R-Square	0.72					

Table 4.10. *Statistics of the fitting with HO subjects (subjects removed: 17, 19, 25, 26, 28, 31).*

HO statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	-0.00122	9.04E-04	-1.234	0.241	-0.003	8.54E-04
B	3.72E-05	1.13E-05	3.286	0.006	1.25E-09	6.18E-05
C	0.00256	9.79E-04	2.613	0.023	4.25E-04	0.005
D	-5.33E-05	1.59E-05	-3.344	0.006	-8.80E-05	-1.86E-05
Number of points	16					
Degrees of Freedom	12					
Reduces Chi-Squared	9.85E-08					
Residual Sum of Squares	1.18E-06					
R Value	0.73					
Adj. R-Square	0.41					

Table 4.11. *Statistics of the fitting with 2B subjects (subject removed: 39).*

2B statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	-0.01163	1.23E-02	-0.945	0.518	-0.168	0.145
B	4.25E-04	3.22E-04	1.321	0.412	-0.004	0.005
C	0.00432	4.67E-03	0.925	0.524	-0.055	0.064
D	-1.28E-04	1.24E-04	-1.034	0.489	0.002	0.001
Number of points	5					
Degrees of Freedom	1					
Reduces Chi-Squared	8.25E-07					
Residual Sum of Squares	8.25E-07					
R Value	0.94					
Adj. R-Square	0.51					

Table 4.12. *Statistics of the fitting with Vicenza subjects (subjects removed: 48, 49).*

Vicenza statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	-0.0034	9.43E-04	-3.898	0.160	-0.0156	8.31E-03
B	8.15E-04	9.43E-05	8.644	0.073	-3.83E-04	2.01E-03
C	1.36E-04	1.32E-04	1.035	0.488	1.54E-03	1.81E-03
D	-5.04E-05	1.34E-05	-3.762	0.165	-2.21E-04	1.20E-04
Number of points	5					
Degrees of Freedom	1					
Reduces Chi-Squared	7.36E-09					
Residual Sum of Squares	7.36E-09					
R Value	0.99					
Adj. R-Square	0.99					

The most significant statistics related to the explicit correlation defined for the proteolytic parameter k_1 are summarized in table 4.13.

Table 4.13. *Summary of goodness of fit statistics.*

	DoF	Reduced chi-sqr	Residuals sum of squares	R^2	\bar{R}^2
HnonO	4	1.11E-08	4.43E-08	0.91714	0.72201
HO	12	9.85E-08	1.18E-06	0.72549	0.40793
2B	1	8.25E-07	8.25E-07	0.93672	0.51018
Vicenza	1	7.36E-09	7.36E-09	0.99951	0.99612

The summary is important to have an immediate visualization of the most important quality-of-fit indexes. As it is clearly visible, the R^2 is high, greater than 90% for almost all the categories, with exception of HO category. This is again due to the high internal variability of the pool of subjects. However, the \bar{R}^2 is relatively high meaning that the correlation can work for the estimation of k_1 with a low error.

4.1.2.1 SGM modification and simulation of MGM_2

The new explicit correlation for k_1 (eq. 4.2) has been substituted in MGM_1, giving rise to MGM_2 (modified Galvanin model, second version). The MGM_2 can be graphically represented as in figure 4.25, where it is clearly visible that both the elimination constant k_e and the proteolytic constant k_1

can be calculated directly from basal clinical trials, whereas k_0 still requires the DDAVP execution to be estimated.

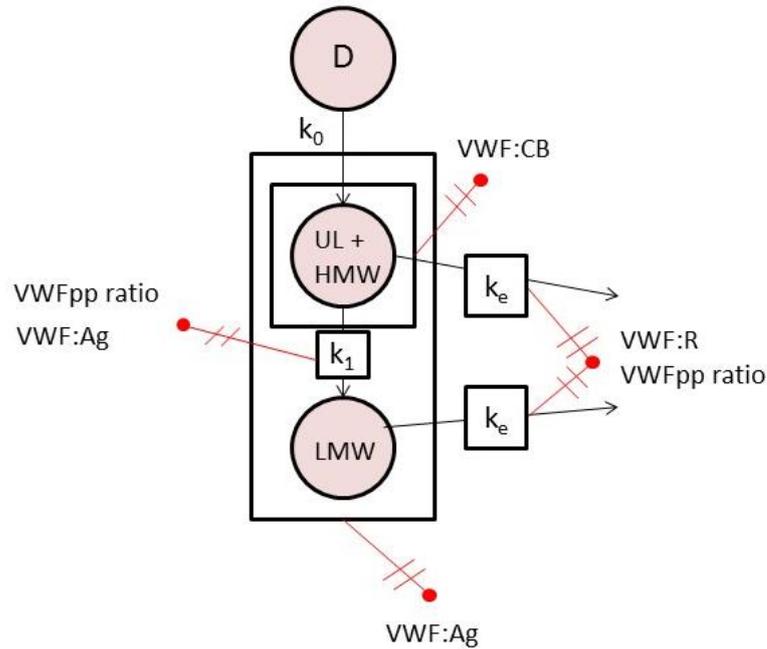


Figure 4.25. *MGM_2 scheme for model identification.*

The algebraic relation introduces four new model parameters (A, B, C, D), whose values have been obtained from the fitting procedure executed in OriginPro[®]. The values of the parameters for each category considered in the study are summarized in table 4.14.

Table 4.14. *Model parameters of the RSM.*

	HnonO	HO	2B	Vicenza
A	-0.00312	-0.00122	-0.01163	-0.00368
B	4.09E-06	3.72E-05	4.25E-04	8.15E-04
C	0.00233	0.00256	0.00432	1.36E-04
<u>D</u>	-2.96E-05	-5.33E-05	-1.28E-04	-5.04E-05

To compare the profiles generated by the SGM (§2.1) and MGM_2, the average subject for each category has been taken as reference. The average input data for each category, required to simulate the models, are those reported in table 4.15.

Table 4.15. *VWFpp ratio and VWF: Ag average values in the considered categories.*

	VWFpp ratio	VWF: Ag [U/dL]
HnonO	1.040	94.07
HO	1.113	77.14
2B	2.320	35.07
Vicenza	0.881	9.42

The comparison between the average k_1 values calculated with the new explicit correlation and the average k_1 estimated values are presented in table 4.16.

Table 4.16. *Calculated average k_1 values and the average estimated k_1 values.*

	k_1 calculated by MGM_1	k_1 estimated through SGM	relative error
HnonO	2.55E-04	2.37E-04	0.0759
HO	4.53E-04	6.25E-04	0.2752
2B	0.00484	0.00471	0.0268
Vicenza	0.00155	0.00149	0.0345

As it is possible to read from table 4.16, the calculated and the estimated values of k_1 are numerically close within each other, indeed the relative error is lower than 10% in all the categories except for HO category. This is due to the high internal variability of the pool of subjects considered in the study for the HO category. This finding reflects on the profiles of VWF:Ag and VWF:CB responses. Figures 4.26, 4.27, 4.28 and 4.29 show the comparison of the profiles of response between the SGM and MGM_2 for all the categories. From the results, it is clearly evident that the profiles of SGM and MGM_2 overlap in each category with exception of HO category, but the deviation ($< \pm 20$ U/dL) produced in the value of the peak of the two model responses VWF:Ag and VWF:CB is still acceptable by the medical community, as found for MGM_1.

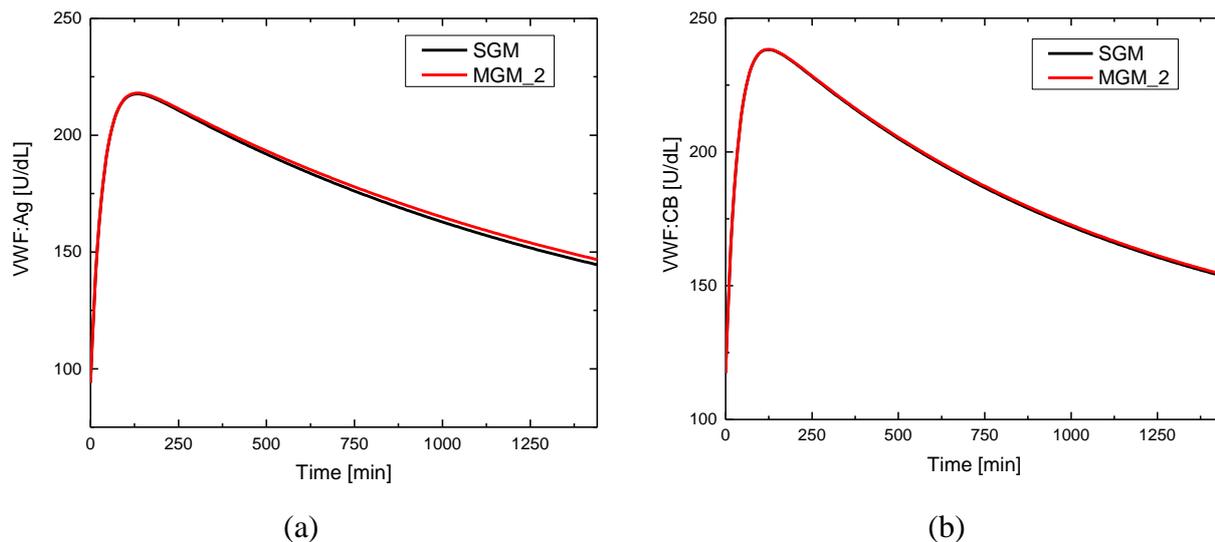


Figure 4.26. Simulated response VWF:Ag (a) and VWF:CB (b) with MGM_2 and SGM in HnonO category.

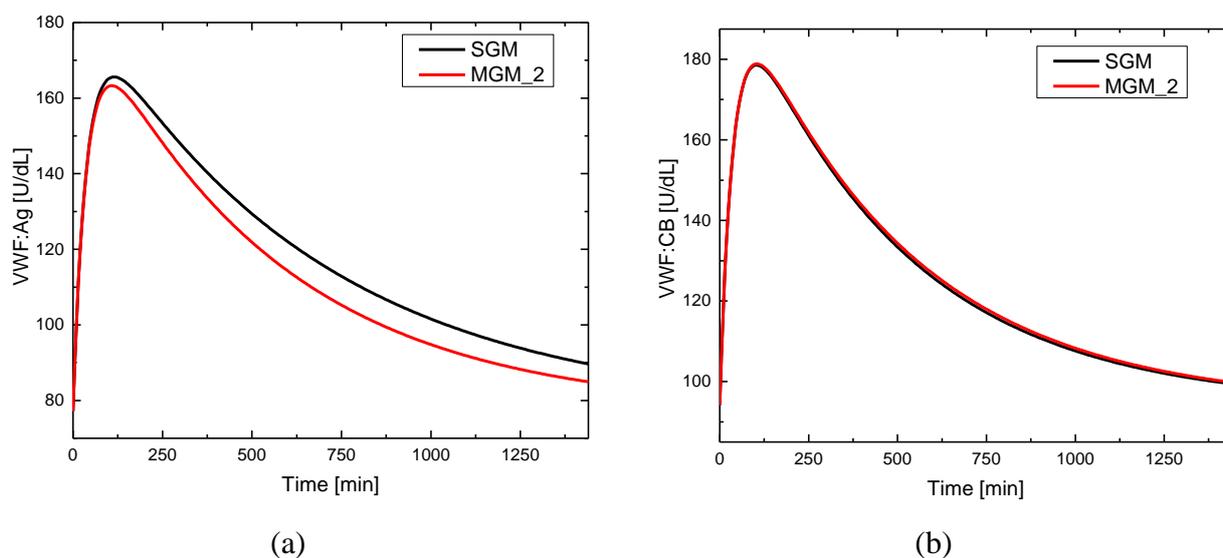


Figure 4.27 Simulated response VWF:Ag (a) and VWF:CB (b) with MGM_2 and SGM in HO category.

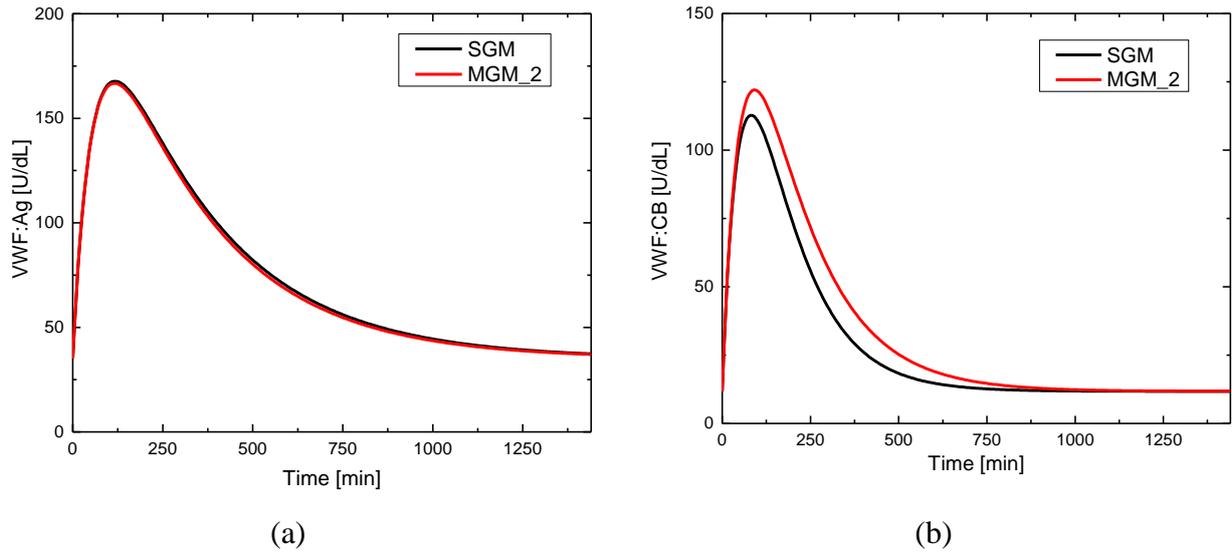


Figure 4.28. Simulated response VWF:Ag (a) and VWF:CB (b) with MGM_2 and SGM in 2B category.

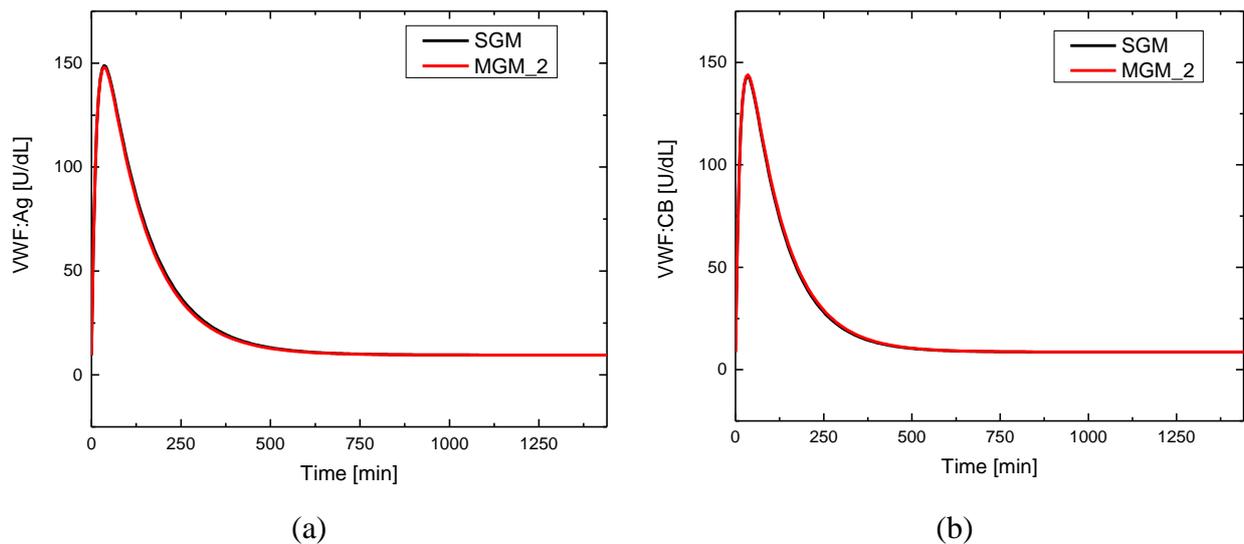


Figure 4.29. Simulated response VWF:Ag (a) and VWF:CB (b) with MGM_2 and SGM in Vicenza category.

4.1.2.2 Information content analysis on MGM_2

Information content analysis has been executed in order to evaluate whether MGM_2 is still locally identifiable.

The local sensitivity analysis has been conducted in gPROMS[®] acting a perturbation of 1% on the release parameter k_0 , which is the last kinetic parameter that requires the DDAVP execution for its estimation.

The sensitivity analysis has been conducted for each category of disease and healthy subjects. Results of the two model responses VWF:Ag and VWF:CB are reported in figures 4.30 and 4.31 for the

healthy subjects HnonO and HO, whereas the profiles of the sensitivity in the 2B and Vicenza categories are presented in figures 4.32 and 4.33, respectively.

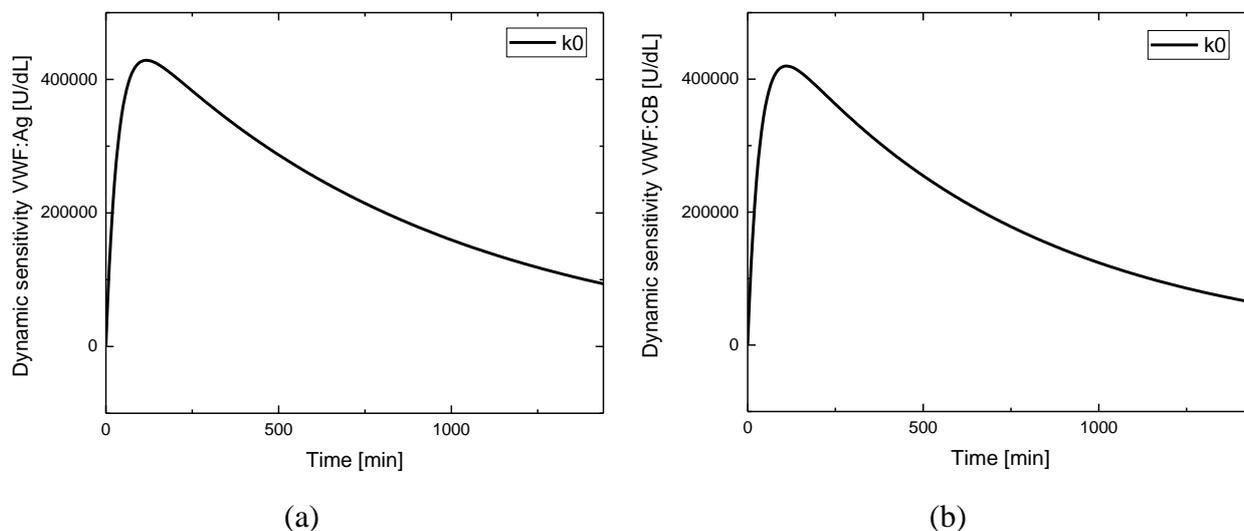


Figure 4.30 Dynamic sensitivity for VWF:Ag (a) and VWF:CB (b) responses in HnonO category.

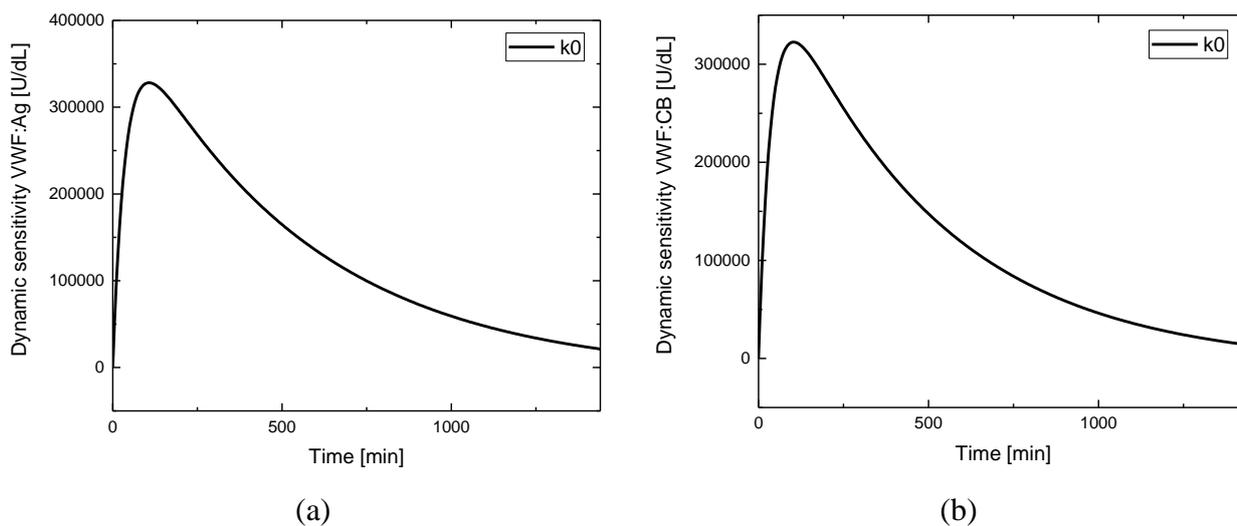


Figure 4.31. Dynamic sensitivity for VWF:Ag (a) and VWF:CB (b) responses in HO category.

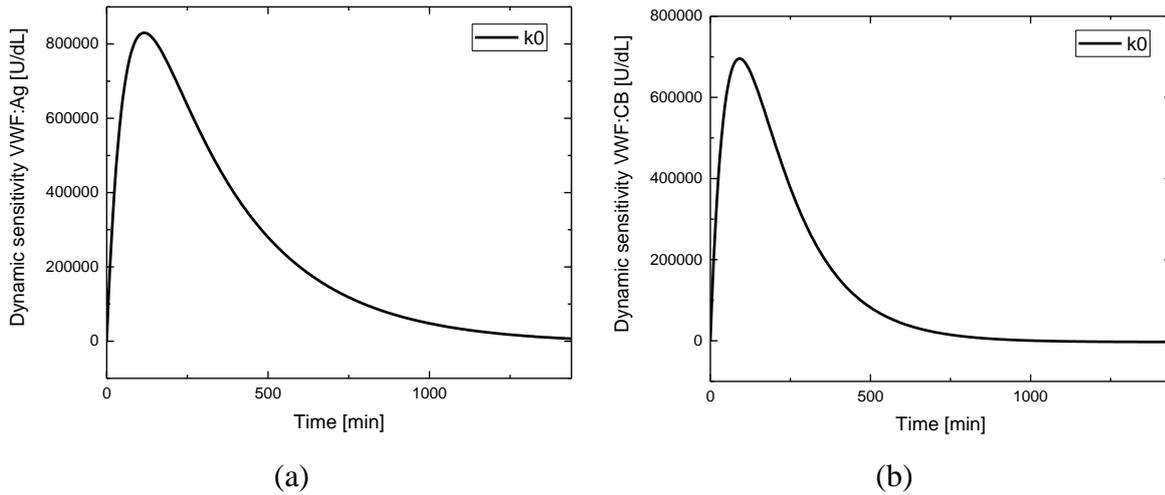


Figure 4.32. Dynamic sensitivity for VWF:Ag (a) and VWF:CB (b) responses in 2B category.

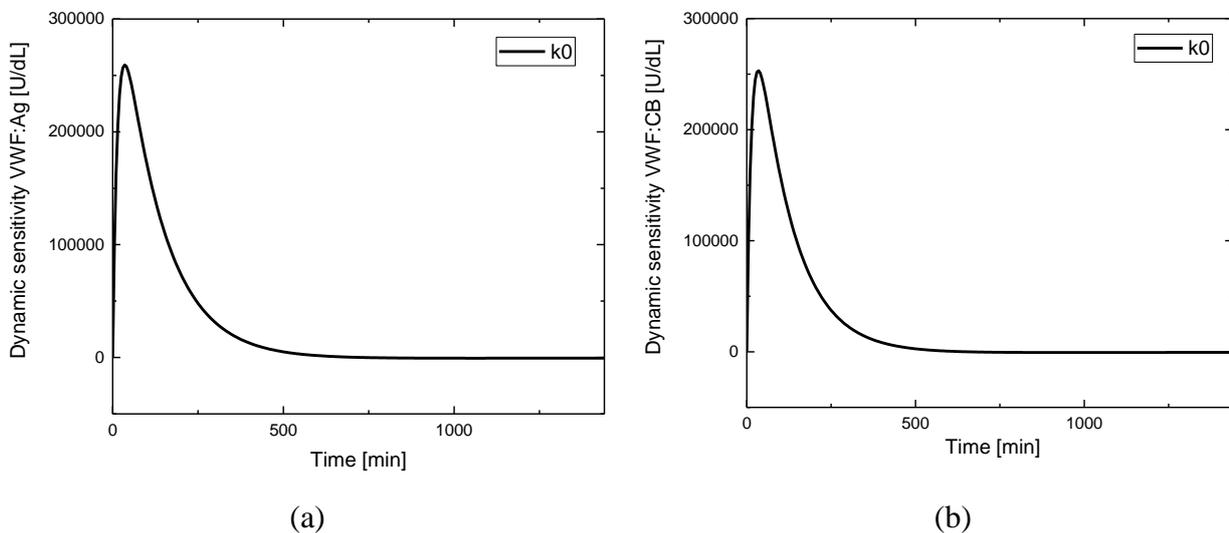


Figure 4.33. Dynamic sensitivity for VWF:Ag (a) and VWF:CB (b) responses in Vicenza category.

As is clearly visible from the results of the sensitivity, the DDAVP execution time can be sensibly reduced to 3-4 hours. Indeed, three-four hours represent a sufficient amount of time for reaching the peak in the sensitivity profiles for all the considered categories. The result is highly promising because the DDAVP execution time might be reduced to 1/6-1/8 of the original test. This will undoubtedly produce a positive impact on the quality of life of the patients. The validation executed in MGM_2 is presented in chapter 5.

4.1.3 RSM for k_0

As has been demonstrated in §4.2, apparently, the use of MGM_2 in VWD characterization allows to reduce significantly the DDAVP execution time. Only the release parameter k_0 requires now the DDAVP execution to be estimated. Therefore, theoretically, if a suitable correlation for parameter k_0 is found, the DDAVP is not necessary anymore and the VWF:Ag and VWF:CB profiles in time of the patients can be simply reproduced using basal quantities values, obtained from a simple blood sample. However, as for k_1 , the research of the right correlation for the release parameter required a big effort (appendix B.3). Indeed, many different clinical quantities can be related to the VWF release path. Furthermore, the release of VWF from the endothelial cells has not been completely understood yet, meaning that there is a lack of knowledge in the physiological description.

For instance, it is impossible to directly evaluate the release rate k_0 . In fact, the basal clinical trials, as the intraplatelet VWF, are able to measure only the amount of VWF released by the cells (Q), and not the rate of release k_0 . However, as soon as the correlation for Q is defined, the kinetic parameter k_0 , can be indirectly obtained from equation 4.3, known D and t_{max} . Precisely, Q is the integral in time of the expression which defines the release physiology in the PK model of VWD (eq. 2.1).

$$Q = \int_0^t k_0 D e^{-k_0(t-t_{max})} \quad (4.3)$$

Thanks to the joint work with the medical school, two correlations have finally been found suitable for the evaluation of Q . The first correlation defines Q as function of VWF:R and intraplatelet VWF (*intra*), which is commonly used by the medical community to evaluate the amount of VWF released by the cells. The mathematical form of the generated linear response surface with interactions is expressed as follow:

$$Q = A + B \cdot \textit{intra} + C \cdot \text{VWF: R} + D \cdot \textit{intra} \cdot \text{VWF: R} \quad (4.4)$$

Fitting the experimental data with the regression model defined in equation 4.4 has given remarkable results, which are presented in figures 4.34, 4.35 and 4.36 for HO, 2B and Vicenza categories. Experimental intraplatelet VWF data were not available for HnonO category.

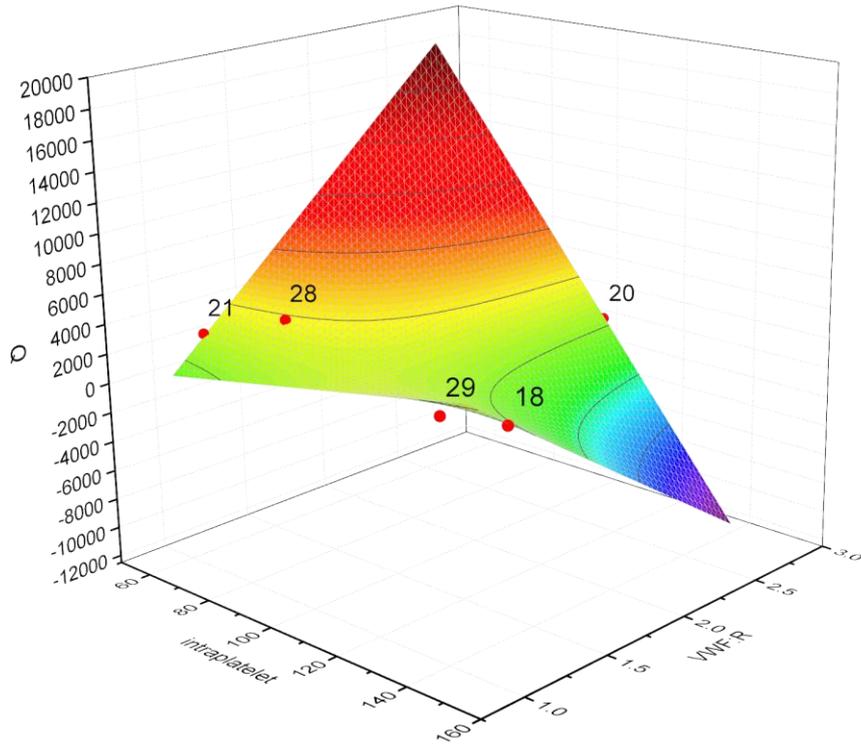


Figure 4.34. Linear response model surface with interactions considering HO subjects (subjects removed: 17, 19, 25, 28, 31).

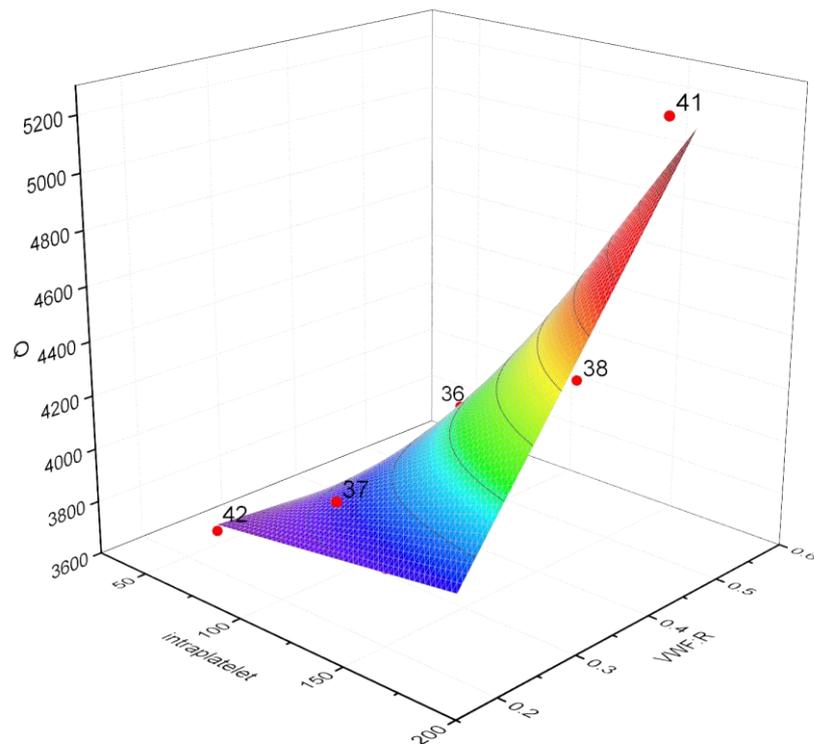


Figure 4.35. Linear response model surface with interactions considering 2B subjects (subject removed: 39).

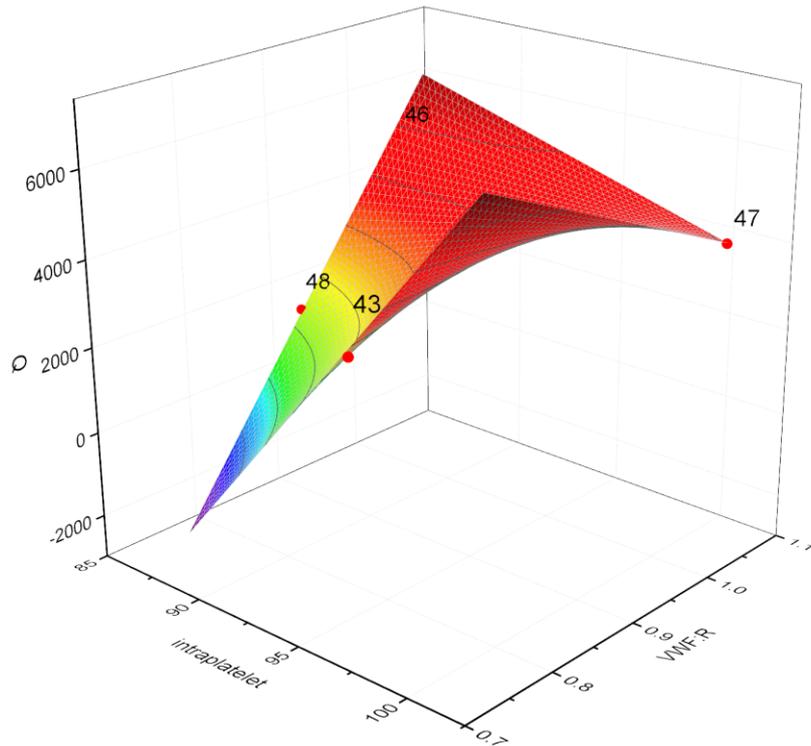


Figure 4.36. Linear response model surface with interactions considering Vicenza subjects (subject removed: 49).

The quality of the fitting is demonstrated also from the values of the statistics reported in tables 4.17, 4.18 and 4.19.

Table 4.17. Statistics of the fitting with HO subjects (subjects removed: 17, 19, 25, 28, 31).

HO statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	-2.25E+04	7.99E+03	-2.819	0.106	-5.69E+04	1.18E+04
B	234.31	71.72	3.267	0.082	-74.27	542.88
C	2.17E+04	6.61E+03	3.277	0.082	-67.84	5.01E+04
D	-203.05	59.78	-3.396	0.077	-460.26	54.17
Number of points	6					
Degrees of Freedom	2					
Reduces Chi-Squared	1.25E+05					
Residual Sum of Squares	2.51E+05					
R Value	0.96					
Adj. R-Square	0.81					

Table 4.18. *Statistics of the fitting with 2B subjects (subject removed: 39).*

2B statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	4.20E+03	1.25E+03	3.349	0.078	-1.2E+03	9.60E+03
B	-7.37	12.09	-0.610	0.604	-59.39	44.64
C	-2.45E+03	3.02E+03	-0.814	0.501	1.54E+03	1.05E+03
D	38.24	27.83	1.374	0.303	-81.48	157.96
Number of points	6					
Degrees of Freedom	2					
Reduces Chi-Squared	1.17E+05					
Residual Sum of Squares	2.34E+05					
R Value	0.92					
Adj. R-Square	0.64					

Table 4.19. *Statistics of the fitting with Vicenza subjects (subjects removed: 49).*

Vicenza statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	-2.71E+05	2.56E+04	-10.580	0.060	-5.96E+05	5.45E+04
B	2.83E+03	266.54	10.609	0.059	-558.87	6.22E+03
C	2.74E+05	2.62E+04	10.471	0.061	-5.84E+04	6.06E+05
D	-2.81E+03	272.15	10.354	0.061	-6.27E+03	641.71
Number of points	5					
Degrees of Freedom	1					
Reduces Chi-Squared	5.88E+04					
Residual Sum of Squares	5.88E+04					
R Value	0.99					
Adj. R-Square	0.97					

As visible from the figures and the statistics just reported, the fitting is good and the correlation could theoretically allow us to satisfy the ambitious target of the research project, which is to completely avoid the execution of the DDAVP clinical trial. The problem is that, again, “simplification” is also one of the goals that must be met in the project. Therefore, correlations must contain basal quantities derived by standard clinical trials. As in the case of VWF:Rco correlation for the proteolytic parameter (see §B.2), here, intraplatelet VWF is a medical test that requires significant effort, good ability in platelets management and uncommon laboratories tools to contain the measurement errors. This means that correlation (4.4) cannot be used at our purpose.

Hence, another correlation has been investigated. The mass balance around the overall control volume of SGM (fig.3) has been executed and, as result, the amount of VWF released is function of the elimination constant k_e and of VWF:Ag. The linear response surface with interactions that has been developed is defined by the following equation:

$$Q = A + B \cdot k_e + C \cdot \text{VWF:Ag} + D \cdot k_e \cdot \text{VWF:Ag} \quad (4.5)$$

The fitting results of the experimental data with the linear response surface (eq. 4.5) are reported in figures 4.37, 4.38, 4.39 and 4.40 for all the considered categories.

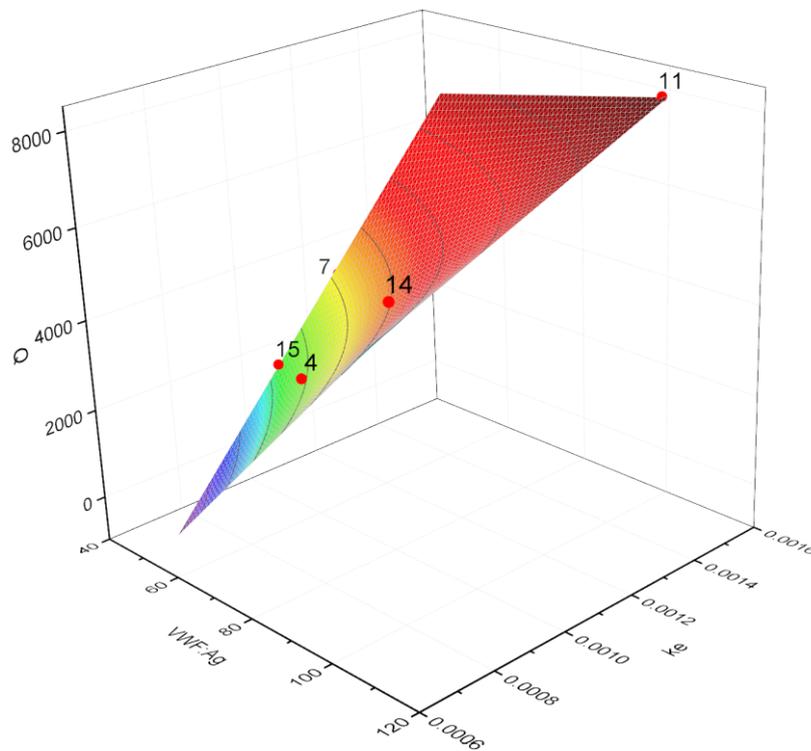


Figure 4.37. Linear response model surface with interactions considering HnonO subjects (subjects removed: 1, 9, 13, 10, 8, 6).

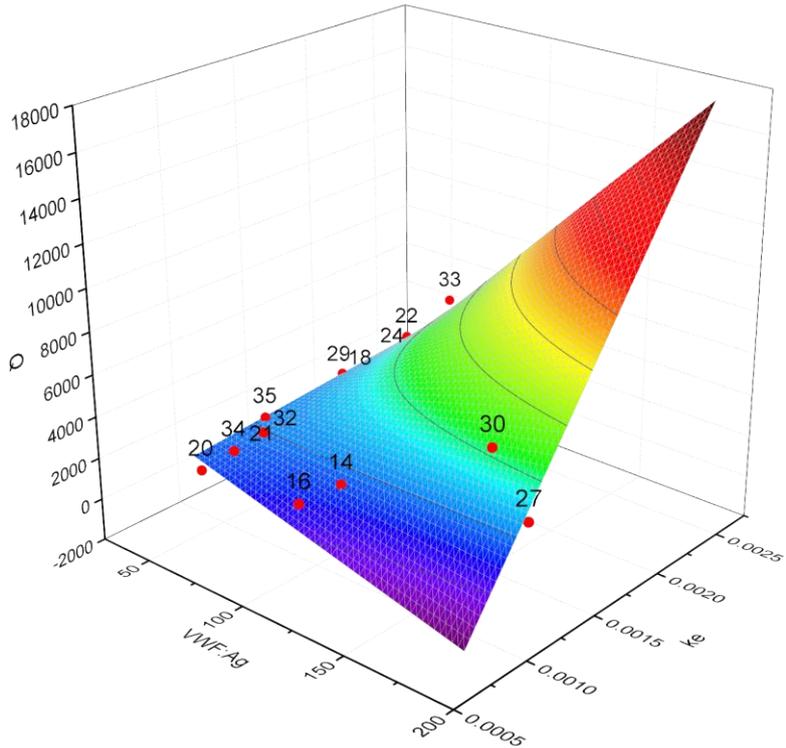


Figure 4.38. Linear response model surface with interactions considering HO subjects (subjects removed: 17, 19, 25, 26, 28, 31).

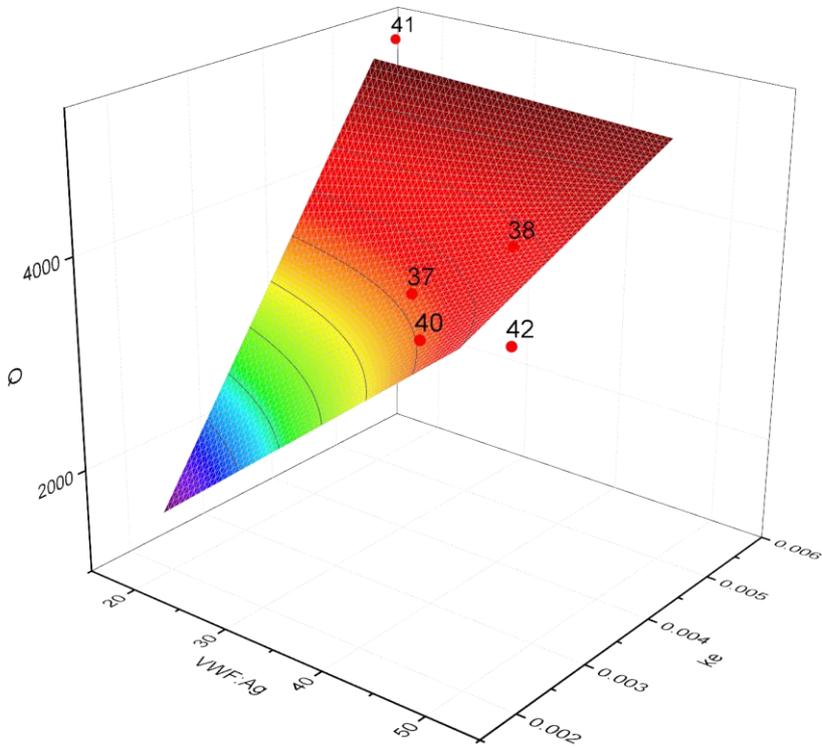


Figure 4.39. Linear response model surface with interactions considering 2B subjects (subject removed: 39).

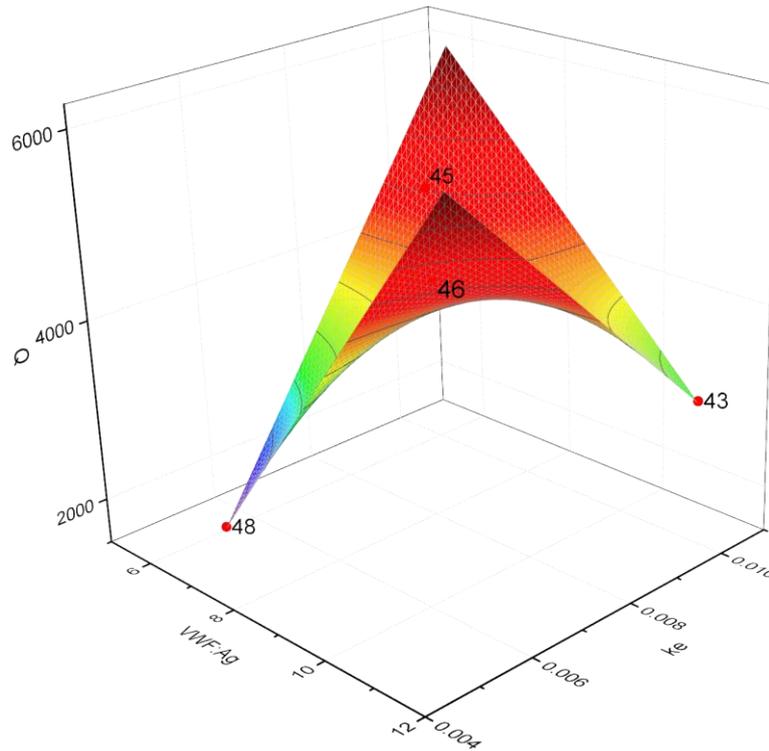


Figure 4.40. Linear response model surface with interactions considering Vicenza subjects (subject removed: 49).

The statistics of the fitting for all the considered categories are reported in tables 4.20, 4.21, 4.22 and 4.23.

Table 4.20. Statistics of the fitting with HnonO subjects (subjects removed: 1, 9, 13, 10, 8, 6).

HnonO statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	-16962.24	10.5E+03	-1.607	0.206	-5.0E+03	1.6E+03
B	191.12	114.53	1.668	0.194	-173.37	555.60
C	1.52E+07	9.72E+06	1.559	0.217	-1.57	4.61
D	-112581.28	10.3E+03	-1.098	0.353	-4.4E+04	2.15E+04
Number of points	7					
Degrees of Freedom	3					
Reduces Chi-Squared	1.12E+06					
Residual Sum of Squares	3.57E+07					
R Value	0.94					
Adj. R-Square	0.75					

Table 4.21 Statistics of the fitting with HO subjects (subjects removed: 17, 19, 25, 26, 28, 31).

HO statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	6.0E+03	1.73E+03	3.470	0.006	241.27	9.8E+03
B	-88.37	26.68	-3.309	0.008	-147.82	-28.91
C	-3.23E+06	1.25	-2.579	0.027	-6.02E+06	-4.4E+05
D	7.55E+03	2.03E+03	-3.778	0.004	3.03E+03	1.2E+05
Number of points	14					
Degrees of Freedom	10					
Reduces Chi-Squared	1.9E+05					
Residual Sum of Squares	1.9E+06					
R Value	0.92					
Adj. R-Square	0.79					

Table 4.22. Statistics of the fitting with 2B subjects (subject removed: 39).

2B statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	-3602.80	4246.15	-0.848	0.458	-1.71E+03	9.91E+03
B	142.16	121.85	1.166	0.327	-245.61	529.94
C	1.64E+06	9.78E+04	1.676	0.192	-1.47E+06	4.75
D	-27636.74	3.14E+04	-0.879	0.444	-1.28E+04	7.23E+04
Number of points	7					
Degrees of Freedom	3					
Reduces Chi-Squared	1.9E+05					
Residual Sum of Squares	5.7E+05					
R Value	0.91					
Adj. R-Square	0.66					

Table 4.23. Statistics of the fitting with Vicenza subjects (subject removed: 49).

Vicenza statistics						
Regression parameters	Value	Standard error	t-Value	Prob> t	95% LCL	95% UCL
A	-18437.71	1445.84	-12.75	0.050	-3.68E+03	-66.51
B	2483.33	166.48	14.92	0.043	368.01	4.60E+03
C	2.72E+06	186.44E+03	14.62	0.043	3.56E+04	5.09
D	-2.98E+05	20.38E+03	-14.66	0.043	5.57E+05	-3.97E+04
Number of points	5					
Degrees of Freedom	1					
Reduces Chi-Squared	3.41E+04					
Residual Sum of Squares	3.41E+04					
R Value	0.99					
Adj. R-Square	0.98					

As it is possible to understand from the results obtained and from the summary of the most important statistics reported in table 4.24 and 4.25 for the response surfaces 4.4 and 4.5, the fitting with the new explicit correlations is good.

Table 4.24. *Summary of goodness of fit statistics for response surface (4.4).*

	DoF	Reduced chi-sqr	Residuals sum of squares	R^2	\bar{R}^2
H0	2	1.25E+05	2.52E+05	0.96	0.81
2B	2	1.17E+05	2.35E+05	0.92	0.64
Vicenza	1	5.88E+04	5.88E+04	0.99	0.97

Table 4.25. *Summary of goodness of fit statistics for response surface (4.5).*

	DoF	Reduced chi-sqr	Residuals sum of squares	R^2	\bar{R}^2
Hnon0	3	1.12E+06	3.57E+07	0.94	0.75
H0	10	1.87E+05	1.87E+06	0.92	0.79
2B	3	1.92E+05	5.74E+05	0.91	0.66
Vicenza	1	3.41E+05	3.41E+05	0.99	0.98

The definition of suitable correlations for the amount of VWF released from the endothelial cells and the consequent evaluation of the release rate k_0 allows us to think that the overall ambitious goal of the project, the DDAVP elimination, may be reached. In the following section, equations (4.4) and (4.5) are singularly tested in MGM_2 developing MGM_3 (modified Galvanin model, third version).

4.1.3.1 SGM modification and simulation of MGM_3

The new correlations for Q have been tested in the MGM_2, developing MGM_3, the third level of modification of SGM. MGM_3 can be represented as in figure 4.41. Precisely, two versions of MGM_3 exist: the first defines Q as in equation 4.4 and it is therefore named as MGM_3_intra, whereas the second evaluates Q through equation 4.5 and it is defined as MGM_3_Ag.

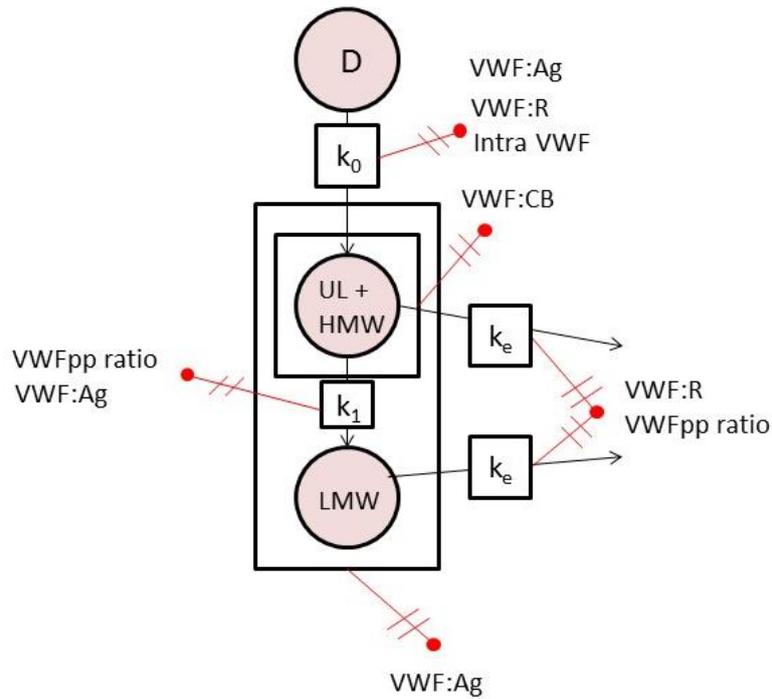


Figure 4.41. MGM₃ scheme.

The algebraic relations introduce four new model parameters (A, B, C, D), whose values have been obtained from the fitting procedure executed in OriginPro[®]. The values of the parameters for each category considered in the study are summarized in table 4.26 and 4.27 for equations (4.4) and (4.5), respectively.

Table 4.26. Model parameters of the RSM for response surface 45.

	HO	2B	Vicenza
A	-22536.65	4202.56	-270557.15
B	234.31	-7.37	2827.81
C	21680.84	-2456.82	274077.65
<u>D</u>	-203.05	38.24	-2816.23

Table 4.27. Model parameters of the RSM for response surface 46.

	HnonO	HO	2B	Vicenza
A	-16962.24	5992.55	-3602.80	-18437.71
B	191.12	-88.37	142.16	2483.33
C	1.52E+07	-3.23E+06	1.64E+06	2.73E+05
<u>D</u>	-112581.28	75479.14	-27636.74	-298690.97

Some subjects belonging to the considered categories (HnonO, HO, 2B and Vicenza) have been taken as reference to compare the profiles generated by the SGM and MGM_3 in both the versions (MGM_3_intra and MGM_3_Ag). The input data required to simulate MGM_3 for every patient are reported in table 4.28.

Table 4.28. *Input data required to simulate MGM_3.*

	Patient 7	Patient 11	Patient 32	Patient 37	Patient 45
	HnonO	HnonO	HO	2B	Vicenza
VWF:Ag [U/dL]	51.70	104.50	63.20	39.60	6.90
VWF:CB [U/dL]	57.20	155.20	93.80	9.30	5.60
VWF:R	1.11	1.48	1.48	0.23	0.81
VWFpp ratio	1.41	1.39	1.39	1.98	6.70
Weight [kg]	60	95	50	76	87

Figures 4.42, 4.43 and 4.44 show the comparison of the profiles of response VWF:Ag between the SGM and MGM_3_intra.

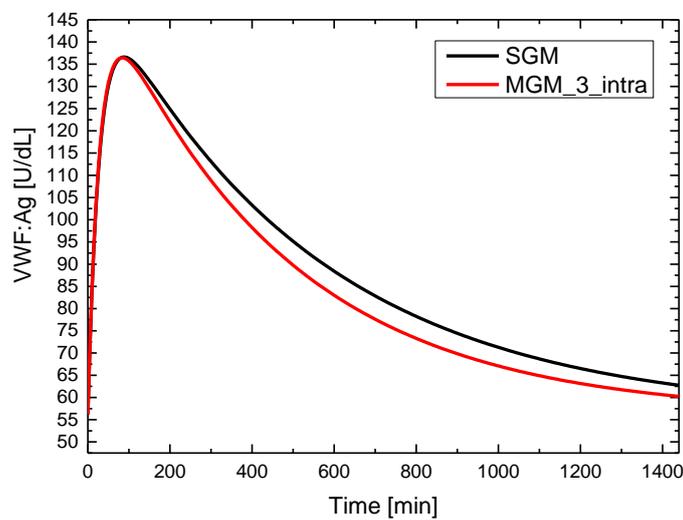


Figure 4.42. *Simulated response VWF:Ag with MGM_3_intra and SGM for patient 32.*

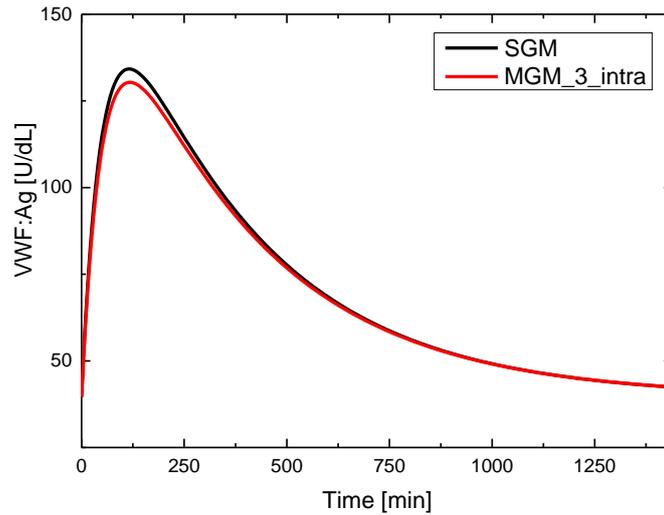


Figure 4.43. Simulated response VWF:Ag with MGM_3_intra and SGM for patient 37.

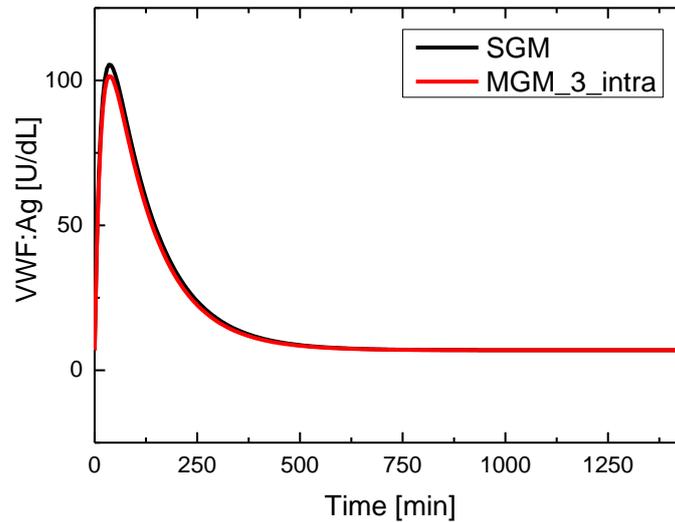
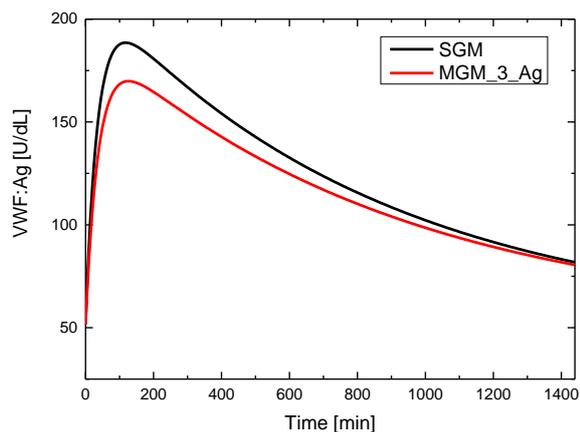


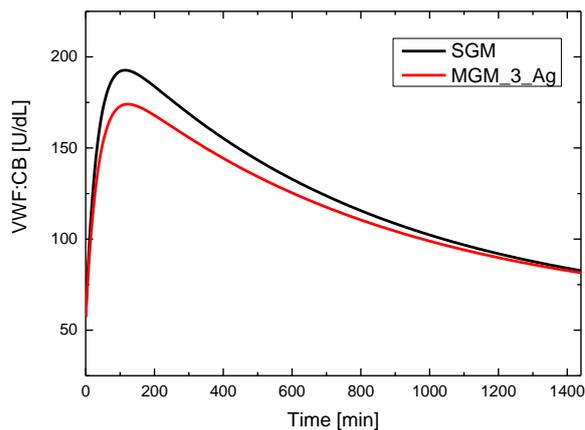
Figure 4.44. Simulated response VWF:Ag with MGM_3_intra and SGM for patient 45.

As figures 4.42, 4.43 and 4.44 illustrates, the MGM_3_intra appears to be able to reproduce the profiles of response for the patients. This means that the correlations (4.1), (4.2) and (4.4) are able to calculate the three kinetic parameters k_0 , k_1 and k_e using only basal clinical values without exploiting the DDAVP. However, as previously stated, correlation (4.4) is reliable, but it implies the execution of the non-standard clinical assay intraplatelet VWF. Therefore, correlation (4.4) should not be used at our purpose, while it is better to test correlation (4.5) in MGM_3.

The comparison of the profiles generated by SGM and MGM_3_Ag are reported in figures 4.45, 4.46, 4.47, 4.48 and 4.49 for all the patients.

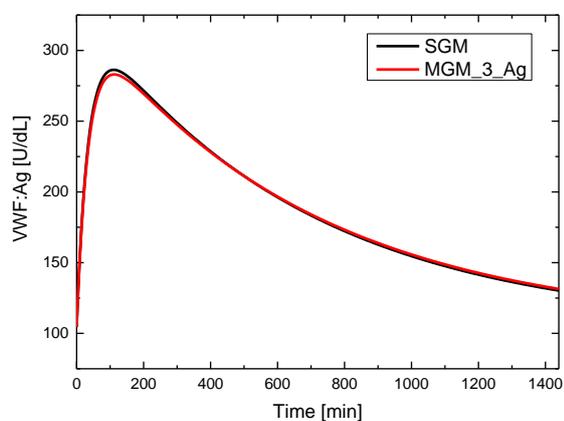


(a)

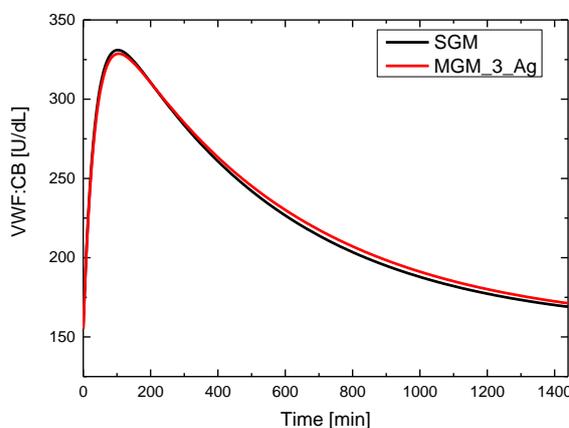


(b)

Figure 4.45. Simulated response VWF:Ag (a) and VWF:CB (b) with MGM_3_Ag and SGM for patient 7.

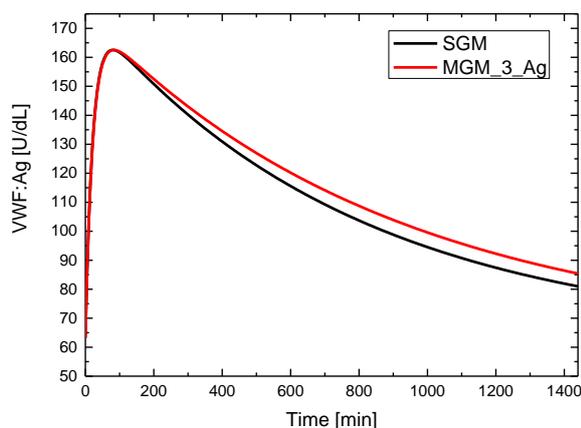


(a)

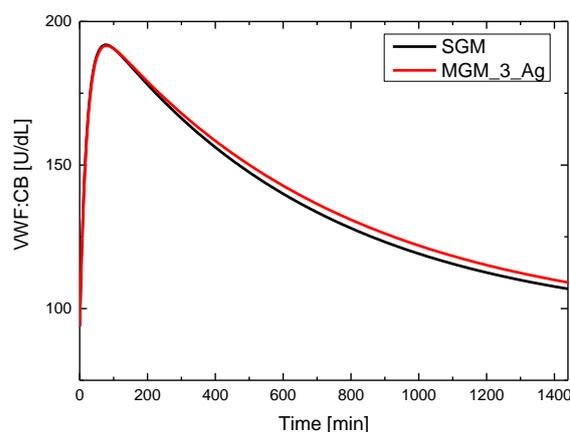


(b)

Figure 4.46. Simulated response VWF:Ag (a) and VWF:CB (b) with MGM_3_Ag and SGM for patient 11.



(a)



(b)

Figure 4.47. Simulated response VWF:Ag (a) and VWF:CB (b) with MGM_3_Ag and SGM for patient 32.

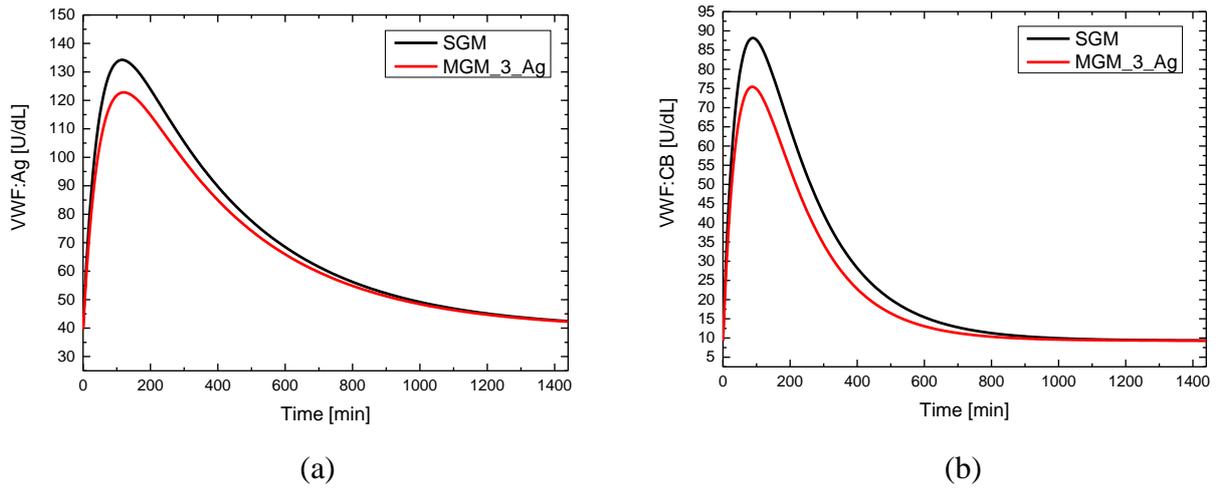


Figure 4.48. Simulated response VWF:Ag (a) and VWF:CB (b) with MGM_3_Ag and SGM for patient 37.

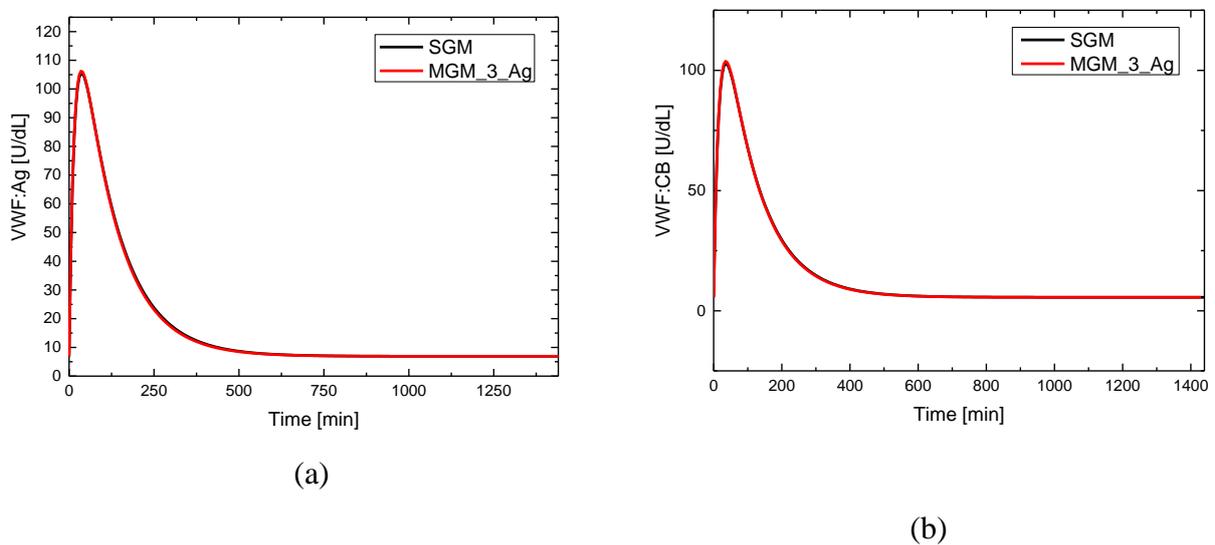


Figure 4.49. Simulated response VWF:Ag (a) and VWF:CB (b) with MGM_3_Ag and SGM for patient 45.

Looking at the results obtained, the MGM_3_Ag simulates quite well the antigen and collagen responses for patients 11, 32 and 45. On the contrary, MGM_3_Ag produces a sensible deviation in the values of the peak for patients 7 and 37 in both the model responses. This is not related to the reliability of the correlations developed for Q , but to the strong assumption made in the model modification procedure. Precisely, suitable correlations have been investigated only for the three kinetic parameters k_0 , k_1 and k_e , while D and t_{\max} , which are the amount and maximum time of VWF release, have been kept constant at the average value for each category. However, as it can be read in table 4.29, the standard deviation from the mean value is large, especially for D .

Table 4.29. Average D and t_{\max} values for each category.

	D [U/dL]	standard deviation	t_{\max} [min]	standard deviation
HnonO	425.41	101.20	75.03	13.36
HO	567.64	154.40	60.03	14.32
2B	597.14	332.47	126.91	42.27
Vicenza	271.29	180.05	46.66	18.21

This means that the correlations (eq. 45 and eq. 46) found for Q are right, but to simulate the profiles of response for a specific subject with precision, the punctual and not the average values for D and t_{\max} are required. In confirmation of this, the profiles of patients 11, 32 and 45 can be correctly simulated because their punctual D and t_{\max} values are closer to the average than those of patients 7 and 37 (table 4.30).

Table 4.30. Punctual D and t_{\max} values for the selected subjects.

Patient	D [U/dL]	t_{\max} [min]
7	558.98	96.11
11	443.06	76.07
32	426.27	43.86
37	294.12	113,14
45	281.82	46.75

Therefore, ideally response surfaces should be developed also for parameters D and t_{\max} . However, this is not possible. Indeed, as can be read in the mathematical definition for Q (eq. 4.4), k_0 , D and t_{\max} are all quantities required in the definition of the amount of VWF released. Moreover, in the article by Galvanin et al. (2017), it was demonstrated that k_0 and D are highly correlated meaning that the two parameters cannot be calculated separately. Furthermore, t_{\max} is not strictly related to any basal clinical quantities but it is function of all the release path, meaning that it is almost impossible to define a basal state correlation for it. In addition, it is important to consider that there is a high degree of uncertainty in the physiological description of the release of VWF in the bloodstream, therefore it is not possible to change the mathematical structure of the differential equations to improve the description of the VWF release. To overcome the described issues, it seems reasonable to execute the DDAVP just for the estimation of the release parameters (D , k_0 and t_{\max}) and then calculate k_1 and k_e using MGM_2. In this way, the main objectives of the research are satisfied.

Indeed, the DDAVP execution time can be sensibly reduced to 3-4 hours, whereas k_1 and k_e correlations are calculated from basal quantities derived from standard clinical trials. However, to confirm what stated, validation has been carried out (see §4.2 for more details).

4.2 Model validation

As described in § 4.1, SGM has been modified into 3 different versions: MGM_1, MGM_2 and MGM_3 (MGM_3_Ag and MGM_3_intra). The main difference between the modified versions can be found in the number of kinetic parameters that are calculated directly from basal state correlations, without considering the DDAVP execution. In particular, MGM_3 distinguishes from the other versions because it evaluates all the kinetic parameters (k_0 , k_1 and k_e) from basal clinical trials (§4.1.3). Therefore, apparently, MGM_3 could help in the achievement of the most ambitious target of the thesis, that is being able to completely eliminate the DDAVP execution for model identification. MGM_3, in both versions, has already been tested trying to simulate the model responses VWF:Ag and VWF:CB of subjects (7, 11, 32, 37, 45) that belong to the pool of the VWD categories considered in this study, but, results were not satisfying (see §4.1.3 for more details). Indeed, MGM_3 is characterized by a big uncertainty in the calculation of the release constant k_0 . The uncertainty is not derived by the mathematical structure of the correlation but by the average values conferred to parameters D and t_{max} . In confirmation to this deduction, model validation has been conducted on MGM_3_Ag, which considers the correlation for k_0 built with standard clinical trials. The subjects taken for validation do not belong to the pool of subjects considered in the study, but to a group of patients that have been removed from the pool because they suffered of some medical issues during the DDAVP execution. Thus, the data derived from an altered DDAVP could not be used in the development of the response surfaces. The required input data of these subjects are reported in table 4.31.

Table 4.31. Required input data for model validation.

Patients	VWF:Ag	VWF:CB	VWFpp ratio	VWF:R	Weight
1	108.9	119.3	0.71	1.10	75
19	58.10	145.20	0.84	0.97	51
39	34.60	12.00	1.65	0.34	50
50	18.10	18.50	11.33	1.02	67

Results are reported in figure 4.50, where it is clearly visible that MGM_3_Ag is not able to correctly describe the VWF release path.

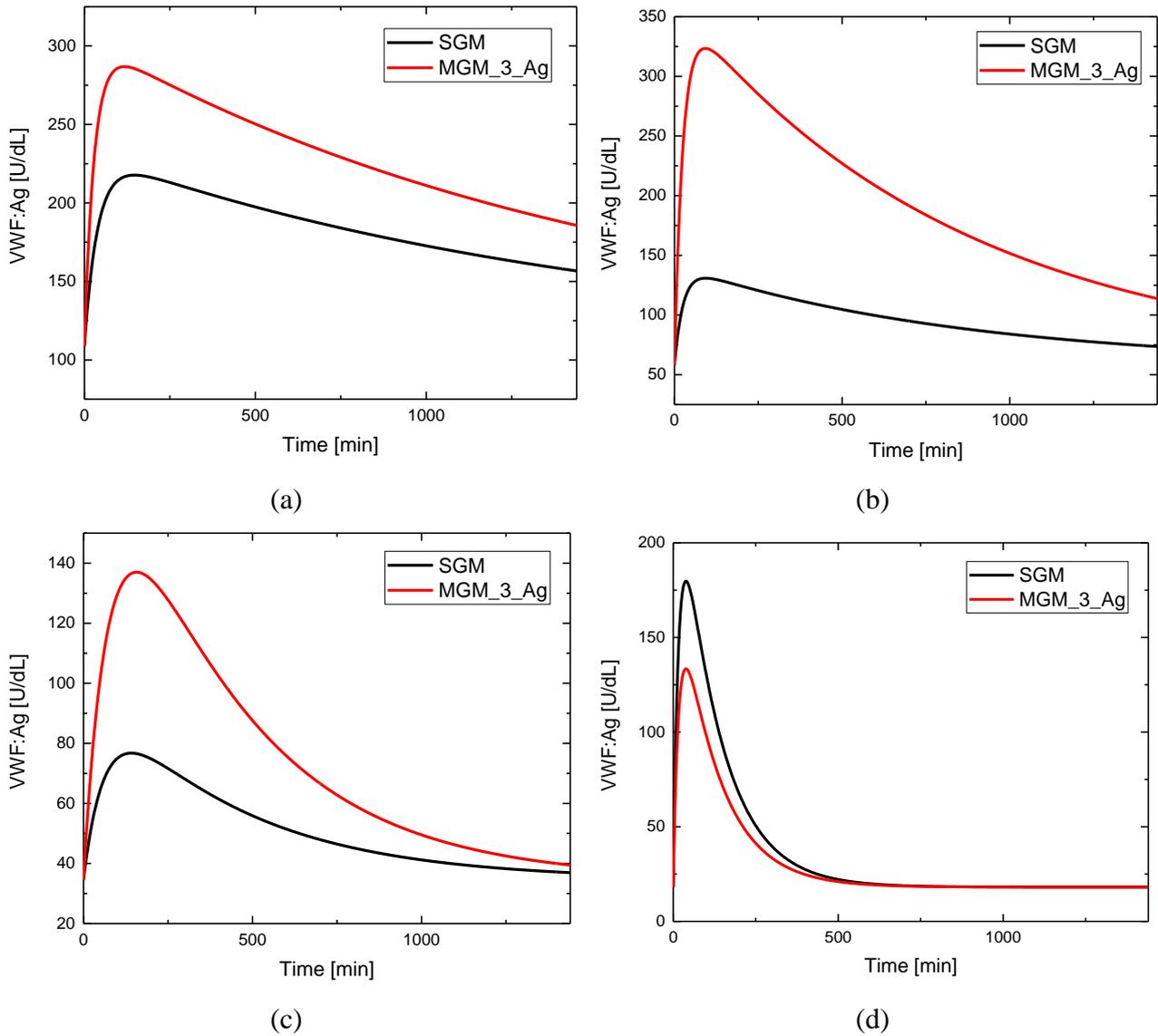


Figure 4.50. Comparison between the simulated VWF:Ag profiles with MGM_3_Ag and SGM for patients 1 (a), 19 (b), 39 (c) and 50 (d).

Hence, with MGM_3_Ag it is not possible to achieve the elimination of the DDAVP execution. Probably, better results could be obtained by using MGM_3_intra, but this correlation has not to be used to respect the important target of simplification, which must be met in our project (see §4.1.3 for more details).

However, it is possible to demonstrate that the main target of the project, which is the reduction in time of the DDAVP execution, can be reached. As seen in §4.1.2, MGM_2 calculates directly from basal clinical trials k_1 and k_e values, whereas k_0 , D and t_{\max} still need to be estimated through DDAVP. Luckily, parameters k_0 , D and t_{\max} are all model kinetic parameters that are required for the description of the VWF release. Therefore, theoretically, the DDAVP execution can be reduced sensibly to the time required for the estimation of the release parameters. Section 4.3 has been entirely dedicated to the redesign of the time-reduced DDAVP. Furthermore, in §4.3, it has been demonstrated that the

simple reduction in time to 3 hours of the 24 hours DDAVP protocol is successful for the estimation of k_0 , D and t_{\max} in all the considered categories.

Model validation has been carried out for MGM_2 with the same subjects used before for the validation of MGM_3. In order to estimate the release parameters k_0 , D and t_{\max} , in silico experimental data have been generated for subjects 1, 19, 39, 50, with a normal random error characterized by zero mean and a variance of four (see appendix C.1 for more details). Results are reported in table 4.32 and they have been used to estimate the model parameters for each subject with the procedure described in §4.3.

Table 4.32. *In silico experimental data for patients 1, 19, 39 and 50.*

Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	108.9	119.3	2.1865	111.0865	121.4865
15	147.06	157.21	2.2185	149.2785	159.4285
30	172.57	182.24	-1.7273	170.8427	180.5127
60	200.78	208.88	0.1547	200.9347	209.0347
120	216.9	221.13	-2.4282	214.4718	218.7018
180	216.94	217.36	-2.227	214.713	215.133

19					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	58.1	145.2	-2.1781	55.9219	143.0219
15	93.63	180.56	0.0651	93.6951	180.6251
30	112.83	199.39	1.1051	113.9351	200.4951
60	127.99	213.52	2.2012	130.1912	215.7212
120	130.02	213.39	3.0884	133.1084	216.4784
180	125.8	207.3	0.1719	125.9719	207.4719

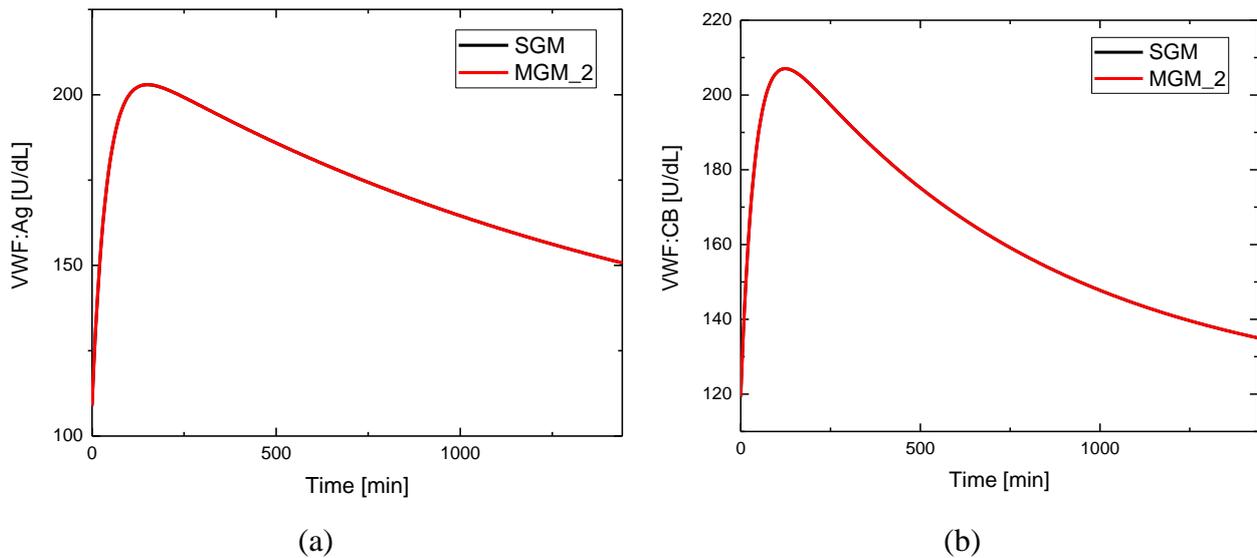
39					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	34.6	12	-0.3848	34.2152	11.6152
15	46.81	23.94	1.7772	48.5872	25.7172
30	56.01	32.44	-1.5297	54.4803	30.9103
60	67.86	42.17	-2.8045	65.0555	39.3655
120	76.33	45.67	-2.8448	73.4852	42.8252
180	75.8	41.07	0.9764	76.7764	42.0464

50					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	18.1	18.5	-1.6089	16.4911	16.8911
15	140.13	140.46	1.3932	141.5232	141.8532
30	176.16	176.35	1.6702	177.8302	178.0202
60	168.16	168.1	-0.4874	167.6726	167.6126
120	113.57	113.3	0.4313	114.0013	113.7313
180	76.2	75.95	-2.3317	73.8683	73.6183

Table 4.33. Results of the estimation of the release parameters with the time-reduced DDAVP for patients 1, 19, 39 and 50.

	1		19		39		50	
	24 h DDAVP	time-reduced DDAVP						
D [U]	470.86	470.30	331.76	349.49	328.31	312.35	421.00	424.67
k_0 [min^{-1}]	0.02570	0.02487	0.04430	0.03832	0.02620	0.02592	0.03690	0.03632
t_{\max} [min]	79.10	75.97	41.32	36.22	80.35	81.12	44.93	46.28

As the results presented in table 5.3 illustrate, the release kinetic parameters can clearly be estimated with precision compared to the values obtained with the 24 hours DDAVP protocol. The new values have been inserted in MGM_2 and the model has been simulated for all the different subjects. Then, results have been compared with the profiles produced with SGM and clearly, MGM_2 is able to reproduce correctly the profiles, once k_0 , D and t_{\max} are estimated with a time-reduced DDAVP. The comparison of VWF:Ag and VWF:CB profiles for patients 1, 19, 39 and 50, is illustrated in figures 4.51, 4.52, 4.53 and 4.54.

**Figure 4.51.** Comparison between the simulated VWF:Ag (a) and VWF:CB (b) profiles with MGM_2 and SGM for patients 1.

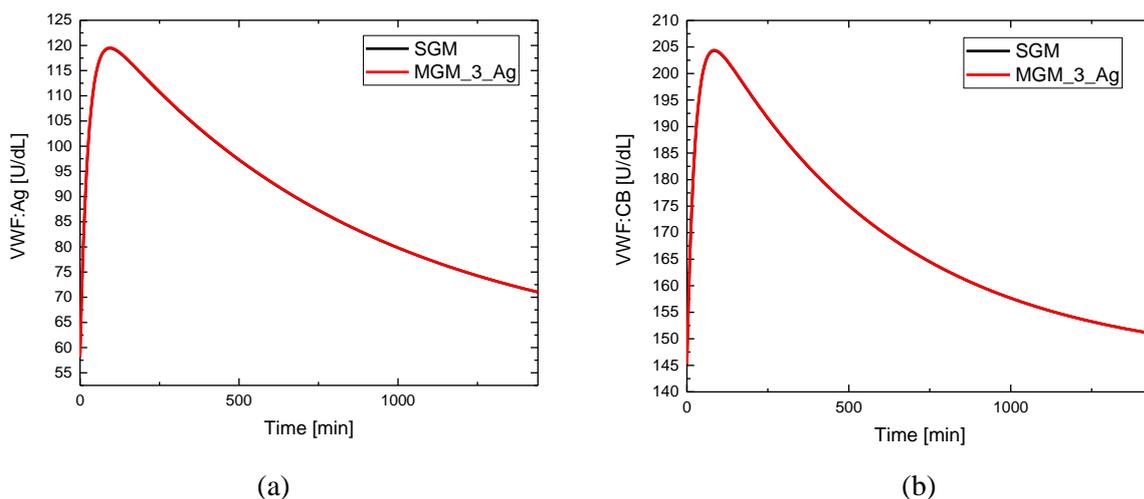


Figure 4.52. Comparison between the simulated VWF:Ag (a) and VWF:CB (b) profiles with MGM_2 and SGM for patients 19.

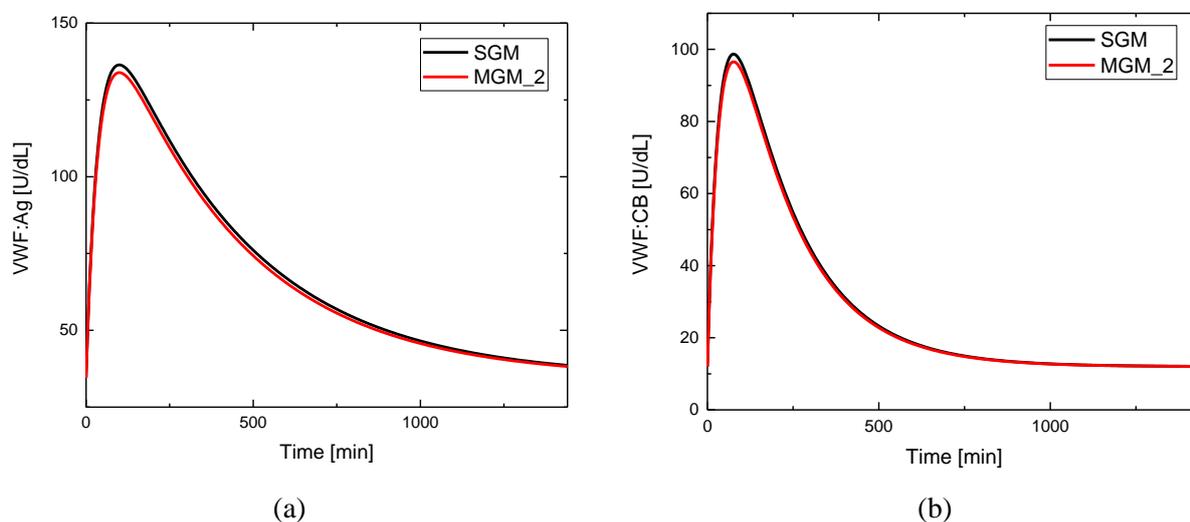


Figure 4.53. Comparison between the simulated VWF:Ag (a) and VWF:CB (b) profiles with MGM_2 and SGM for patients 39.

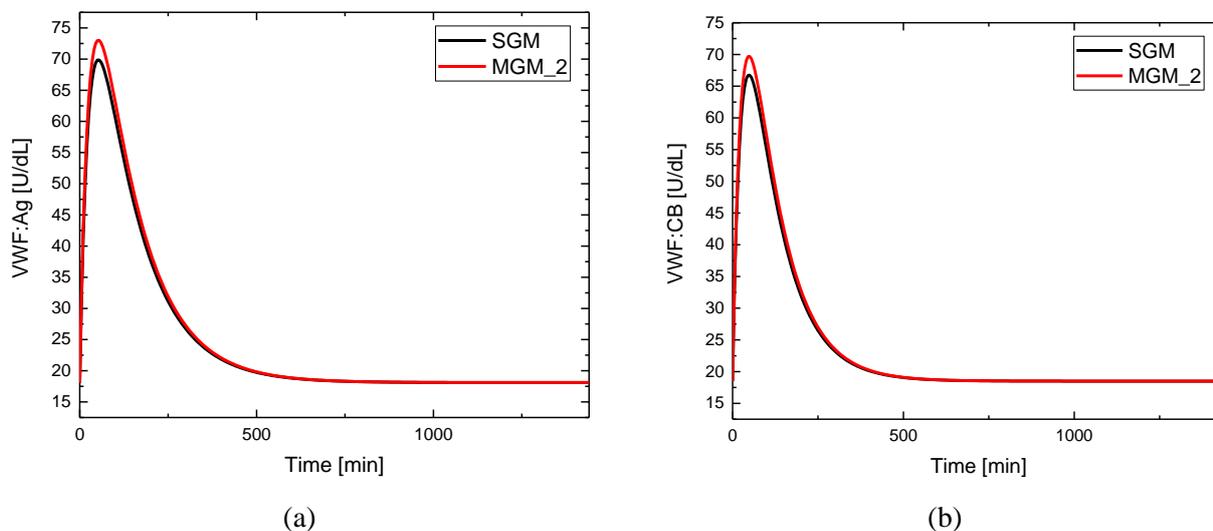


Figure 4.54. Comparison between the simulated VWF:Ag (a) and VWF:CB (b) profiles with MGM_2 and SGM for patients 50.

Moreover, table 4.33 shows the comparison between the values of the kinetic parameters (k_0 , D , t_{\max} , k_1 and k_e) obtained by the combination of MGM_2 and the time-reduced DDAVP and the values of the parameters estimated through the original 24 hours DDAVP protocol with SGM.

Table 4.33. Comparison of the model parameters of subjects 1, 19, 39 and 50.

	1		19		39		50	
	24 h DDAVP	MGM_2 + time-reduced DDAVP	24 h DDAVP	MGM_2 + time-reduced DDAVP	24 h DDAVP	MGM_2 + time-reduced DDAVP	24 h DDAVP	MGM_2 + time-reduced DDAVP
k_0 [min^{-1}]	0.0257000	0.0248700	0.04431520	0.03832008	0.02620	0.025921	0.03690	0.03632
k_1 [min^{-1}]	0.0006890	0.0006995	0.00056854	0.00057122	0.002875	0.0028663	0.00232	0.0022759
k_e [min^{-1}]	0.0006520	0.0006489	0.00118132	0.00118036	0.002671	0.0025464	0.00831	0.0083133
D [U]	470.86	470.30	331.76	349.49	328.31	312.35	421.00	424.67
t_{\max} [min]	79.10	75.97	41.32	36.22	80.35	81.12	44.93	46.28

In conclusion, as results demonstrate, the approach for model identification proposed in this thesis, that is, combining MGM_2 and the time-reduced DDAVP, allows to successfully estimate the kinetic parameters of subjects that belong to the considered categories (HnonO, HO, 2B and Vicenza), but that are not part of the pool of subjects used to define the response surfaces. Thus, this means that validation gives positive results.

4.3 Redesign of the DDAVP clinical trial

As illustrated in section 4.1.3, the VWF release was found critical to define. Indeed, the linear response surface with interactions may work accurately in the calculation of the release kinetic constant, but D and t_{\max} cannot be used at the average value of each category. Therefore, it seems reasonable to exploit MGM_2 (§4.2) and redesign the DDAVP clinical trial to estimate the release parameters. For instance, to redesign the test, the trace of FIM in MGM_2 has been evaluated. In this case, the trace evaluates the information content that can be brought by the parameters that define the release k_0 , D and t_{\max} (figure 4.55.b).

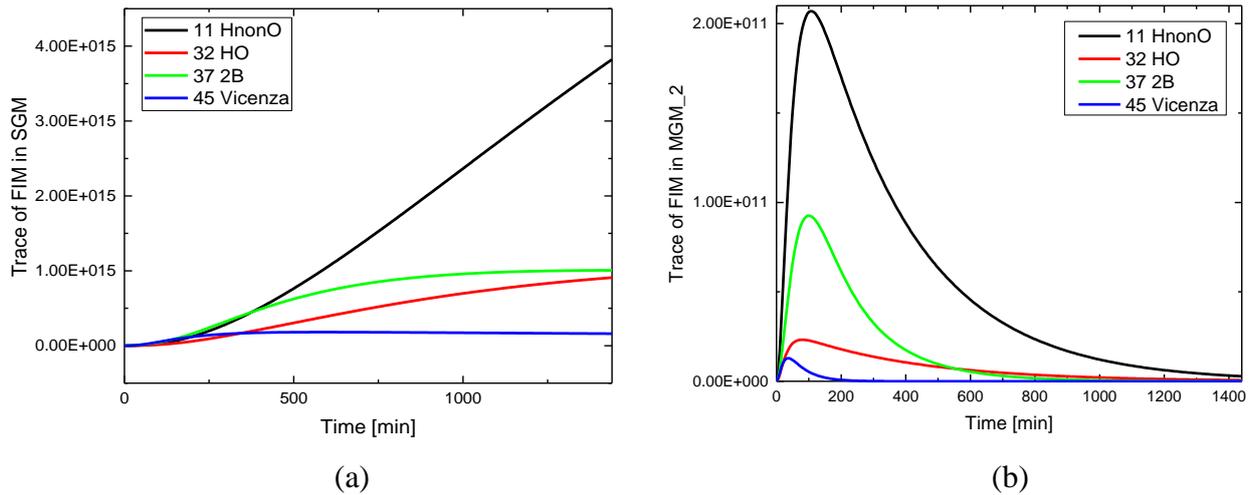


Figure 4.55. Trace of FIM considering all the VWD categories in SGM (a) and MGM_2 (b).

As figure 4.55.b demonstrates, three hours appear to be a sufficient amount of time in MGM_2 for reaching the peak in the information content considering all the VWD categories. On the contrary, the maximum information content cannot be reached by SGM in the DDAVP execution time, as visible in figure 4.55.a. Hence, considering the results obtained from the analysis of the information content, it may be possible to exploit MGM_2 and reduce the DDAVP to 3 hours.

Initially, the original 24 h DDAVP design [0, 15, 30, 60, 120, 180, 480, 1440] has been simply reduced.

The reduction acts both in the execution time and in the number of sampling points. Indeed, the time-reduced DDAVP [0, 15, 30, 60, 120, 180] lasts for 3 hours instead of 24 hours and the number of sampling points is now fixed to 6 instead of 8.

The time-reduced DDAVP design required validation to understand if the release parameters k_0 , D and t_{\max} can still be estimated with precision. Thus, a random set of experimental data has been generated for each category, considering a normal random error with zero mean and a variance of 4. The set of experimental data used for design validation are reported in appendix C (§C.1). The patients taken as reference to conduct the validation of the time-reduced DDAVP are: patient 11 for HnonO, 32 for HO, 37 for 2B and 45 for Vicenza categories. In silico experimental data have been generated to perform parameter estimation, whose applied procedure, based on each subject's VWF:Ag and VWF:CB readings, is:

Step 0: all parameters k_0 , D and t_{\max} are left free to vary starting from the initialization value that is the punctual value for each subject estimated through the 24 hours DDAVP design with SGM.

Step 1: t_{\max} is set at the value used in the previous step, while k_0 and D are estimated.

Step 2: k_0 and D are set at the value used in the previous step, while t_{\max} is estimated.

Step 1 and step 2 are repeated until the estimates do not vary significantly (i.e. until the difference between the estimates is lower than 0.1 % for each parameter).

Results of the estimation and the comparison between the 24 hours DDAVP and the time-reduced DDAVP are reported in table 4.34.

Table 4.34. Parameter estimation results (patients 11, 32, 37 and 45) executed with the time-reduced DDAVP ($t_{ref} = 0.1837$).

		11				
		24 h DDAVP	time-reduced DDAVP	95% conf. Interval	95% t-value	standard deviation
D	[U]	558.98	578.35	120	4.6	56
k_0	[min ⁻¹]	0.0279	0.0282	0.0021	13	0.00094
t_{\max}	[min]	96.11	93.75	0.63	150	0.29

		32				
		24 h DDAVP	time-reduced DDAVP	95% conf. Interval	95% t-value	standard deviation
D	[U]	294.12	293.18	93	3.2	42
k_0	[min ⁻¹]	0.04585	0.04586	0.0066	7	0.0029
t_{\max}	[min]	43.86	43.93	0.71	62	0.32

		37				
		24 h DDAVP	time-reduced DDAVP	95% conf. Interval	95% t-value	standard deviation
D	[U]	426.27	426.32	160	2.6	72
k_0	[min ⁻¹]	0.0197	0.0197	0.0029	6.8	0.0013
t_{\max}	[min]	113.15	113.14	2.1	54	0.96

		45				
		24 h DDAVP	time-reduced DDAVP	95% conf. Interval	95% t-value	standard deviation
D	[U]	281.81	280.33	120	2.2	56
k_0	[min ⁻¹]	0.0606	0.0607	0.0088	6.9	0.0039
t_{\max}	[min]	46.76	46.86	0.63	74	0.29

As it is possible to understand from the values of the parameters obtained, the time-reduced design is able to correctly estimate the release parameters of each patient. The standard deviation for parameter D is quite high, but, the goodness of the estimation is then reinforced by the t-values, which are substantial for all the subjects in the pool.

However, the design used to get the experimental data for estimating the release parameters is just a reduction in time of the original 24 h DDAVP protocol. Thus, the time-reduced design has to be optimized for a reduced time-horizon. To optimize the time-reduced DDAVP protocol, the design of experiment package in gPROMS[®] has been used. The applied procedure can be summarized as follow:

Step 0: selection of the subjects of reference;

Step 1: sampling time optimization in gPROMS[®] for each subject;

Step 2: new DDAVP protocol definition;

Step 3: new DDAVP protocol validation.

Four subjects for each of the considered categories (HnonO, HO, 2B and Vicenza) have been taken as reference to conduct the optimization. The selection of the patients was executed considering subjects with sensibly different values of the model parameters k_0 , D and t_{\max} even if belonging to the same class. This approach has been chosen to guarantee a complete description of the subjects in the pool. The patients taken as reference for each category are reported in table 4.35.

Table 4.35. *Selected subjects.*

	Patient			
HnonO	11	5	15	14
HO	32	20	30	18
2B	37	38	41	40
Vicenza	45	48	43	46

Punctual data of each subject (appendix A) have been inserted in gPROMS[®] and the time-reduced DDAVP [0, 15, 30, 60, 120, 180] has been used as initial guess for the sampling time to optimize. Then, the optimization has been carried out considering three experiment design approaches: A-optimal, D-optimal and E-optimal. Results of the punctual optimization executed for each selected subject are illustrated in appendix C (§C.2).

The DDAVP protocol must be unique for all the categories. Indeed, theoretically, the class of belonging of the patients is at the beginning unknown. To unify the protocol, the average value of each optimized sampling point for all the selected subjects among the categories has been evaluated. Results are reported in table 4.36 for the three designs.

Table 4.36 *Average optimized sampling time among the categories.*

Average A	Round A	Standard	Average D	Round D	Standard	Average E	Round E	Standard
[min]	[min]	deviation	[min]	[min]	deviation	[min]	[min]	deviation
0	0	0	0	0	0	0	0	0
18.27	18	6.4	20.83	20	8.0	18.27	18	6.4
36.51	36	9.9	37.32	35	9.0	36.51	36	9.9
61.49	60	11.6	55.39	55	10.3	61.49	60	11.5
156.84	155	42.9	141.09	140	44.5	156.84	155	42.9
171.83	172	42.9	165.98	166	46.5	171.85	172	43.0

As is visible in the results obtained, A and E optimal designs produce an identical optimized sampling time considering the average values, therefore just two optimized designs have been considered (A-optimal and D-optimal). Furthermore, the standard deviation sensibly increases in the last two sampling points. This is linked to the DDAVP profile that behaves differently among the categories. In fact, the time required for reaching the peak of release in Vicenza category is almost half of the time required by HnonO subjects. On the contrary, the standard deviation is lower in the initial sampling points because at the beginning the DDAVP trend is similar between the categories.

A-optimal and D-optimal designs have then been tested and validated. To test their robustness, the designs have been carried out for the subjects that are characterized by the highest variance-covariance matrix. In table 4.36, the determinant of the variance-covariance matrix has been reported, for each subject considered, as metric to define the most critical situations. This approach is called “worst-case” method and it represents the most intuitive way for evaluating the robustness of a specific design. The patients, which have been considered for validation are:

- Patient 5 for HnonO;
- Patient 20 for HO;
- Patient 38 for 2B;
- Patient 48 for Vicenza.

Table 4.36. Determinant of the variance-covariance matrix for the different subjects taken as reference.

	Patient			
HnonO	11	5	15	14
Determinant	5.68E-03	4.1E-02	7.30E-03	5.56E-03
HO	32	20	30	18
Determinant	5.34E-02	5.20E-01	3.9E-02	7.51E-02
2B	37	38	41	40
Determinant	2.19E-02	1.05E-01	1.22E-02	9.24E-03
Vicenza	45	48	43	46
Determinant	1.12E-01	2.27	3.56E-01	5.06E-01

Parameter estimation has been executed for all the subjects taken as reference (5, 20, 38, 48), following the procedure applied in the time reduced DDAVP design. Results are reported in table 4.37 for A-optimal design and in table 4.38 for D-optimal design.

Table 4.37. *Parameter estimation through A-optimal protocol.*

		5				
		24 h DDAVP	A-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	407.02	385.64	140	2.8	62
k_0	[min-1]	0.0294	0.0299	0.0046	6.5	0.0021
t_{max}	[min]	69.19	69.56	1.2	60	0.53

		20				
		24 h DDAVP	A-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	412.17	453.08	220	2	100
k_0	[min-1]	0.0313	0.0298	0.0099	3	0.0044
t_{max}	[min]	40.1	40.86	3.2	13	1.4

		38				
		24 h DDAVP	A-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	930.25	583.59	200	2.9	89
k_0	[min-1]	0.0086	0.0168	0.0025	6.6	0.0011
t_{max}	[min]	176.32	109.7	2.2	49	1

		48				
		24 h DDAVP	A-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	50.6	47.43	120.00	0.41	52
k_0	[min-1]	0.0996	0.0993	0.068	1.5	0.031
t_{max}	[min]	34.54	33.79	1.2	28	0.54

Table 4.38. *Parameter estimation through D-optimal protocol.*

		5				
		24 h DDAVP	D-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	407.02	342.23	120	2.7	56
k_0	[min-1]	0.0294	0.0322	0.0048	6.7	0.0021
t_{max}	[min]	69.19	71.06	1.1	64	0.51

		20				
		24 h DDAVP	D-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	412.17	441.789	220	2	100
k_0	[min-1]	0.0313	0.0298	0.01	3	0.0045
t_{max}	[min]	40.9	39.95	3.2	12	1.5

		38				
		24 h DDAVP	D-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	930.25	496.39	180	2.7	82
k_0	[min-1]	0.00855	0.01792	0.0028	6.5	0.0012
t_{max}	[min]	176.32	109.02	2.2	49	1

		48				
		24 h DDAVP	D-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	50.6	50.58	130	0.39	58
k_0	[min-1]	0.0996	0.09906	0.072	1.4	0.032
t_{max}	[min]	34.54	34.34	1.1	31	0.5

As is clearly visible from the results obtained, the standard deviation is high for parameter D in all the selected subjects, but the t-value is generally not bad. As can be read from the tables, the values estimated by the new designs with MGM_2 are close to the values estimated by SGM, even if the sampling points are different. The results can be better visualized in figures 4.56 and 4.57.

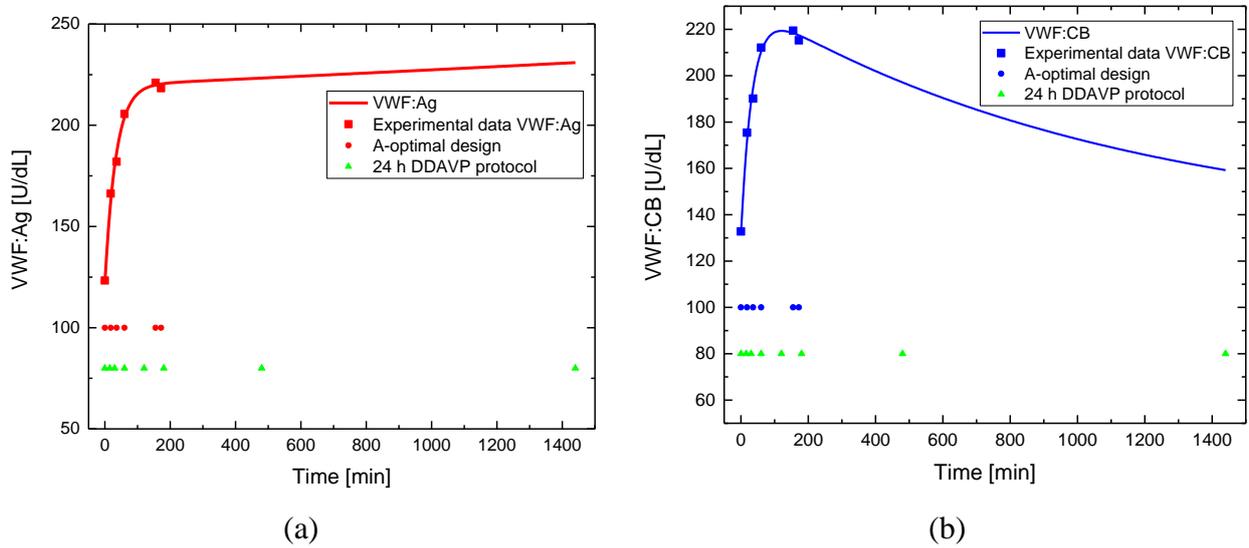


Figure 4.56. Comparison between the generated and predicted VWF:Ag (a) and VWF:CB (b) profiles for patient 5 using the A-optimal design.

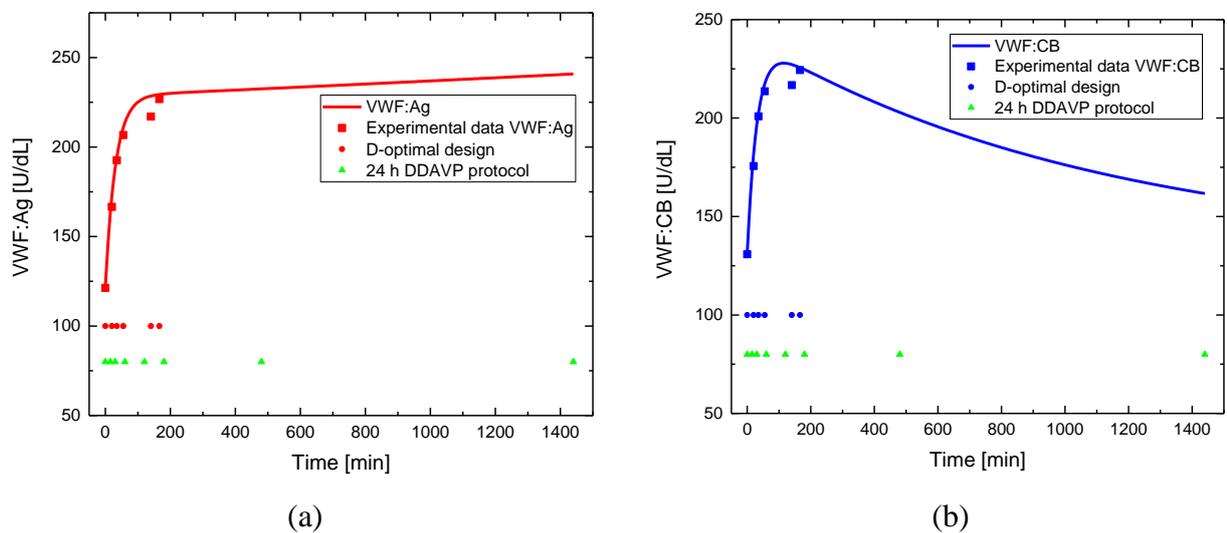


Figure 4.57. Comparison between the generated and predicted VWF:Ag (a) and VWF:CB (b) profiles for patient 5 using the D-optimal design.

Figure 4.56 illustrates the comparison between the experimental data and the predicted profiles for VWF:Ag (a) and VWF:CB (b) responses for patient 5, using the parameters estimated with the A-optimal design. The same comparison is shown in figure 4.57, in which, differently, the parameters have been estimated with the D-optimal design. As is clearly visible, both the optimized time-reduced

designs A-optimal and D-optimal allow to estimate correctly the parameters. Indeed, the predicted profiles approximate well the generated experimental data.

To test the robustness of the redesigned DDAVP clinical trial, the parameter estimation procedure has been again executed on the same subjects (5, 20, 38, 48) following the A-optimal and D-optimal sampling times, but based on a set of experimental data characterized by an error with a variance of 36. This analysis is quite important, because in the clinical procedures the error made on measurements execution is usually not negligible and the designs must be sufficiently robust to deal with that.

The generate experimental data are reported in table C.3 of appendix C (§C.1). Results are reported in table 4.39 and 4.40 for A-optimal and D-optimal designs, respectively. As tables illustrate, the standard deviations associated to parameter D are higher and the t-values are sensibly lower than those calculated with the experimental data set generated with a variance of 2. However, the estimation of the parameters k_0 , D and t_{max} is acceptable, even if the related uncertainty increases significantly.

Table 4.39. Parameter estimation through A-optimal protocol ($t_{ref} = 0.1534$).

		5				
		24 h DDAVP	A-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	407.02	405.24	1300	0.32	570
k_0	[min-1]	0.02934	0.02919	0.041	0.71	0.018
t_{max}	[min]	69.19	69.19	11	6.5	4.8

		20				
		24 h DDAVP	A-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	412.17	416.93	1900	0.21	870
k_0	[min-1]	0.0313	0.0321	0.096	0.34	0.043
t_{max}	[min]	40.09	41.92	28	1.5	13

		38				
		24 h DDAVP	A-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	930.25	467.17	1500	0.31	680
k_0	[min-1]	0.00855	0.0184	0.025	0.74	0.011
t_{max}	[min]	176.32	109.2	20	5.4	9.2

		48				
		24 h DDAVP	A-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	50.6	47.8	1000	0.046	460
k_0	[min-1]	0.0996	0.0996	0.61	0.16	0.27
t_{max}	[min]	34.54	34.54	9.9	3.5	4.5

Table 4.40. Parameter estimation through D-optimal protocol.

		5				
		24 h DDAVP	D-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	407,015	415.83	1200	0.35	540
k_0	[min ⁻¹]	0.0294	0.0307	0.037	0.82	0.017
t_{\max}	[min]	69.19	73.4	9.3	7.9	4.2

		20				
		24 h DDAVP	D-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	412.17	419.7	1900	0.22	860
k_0	[min ⁻¹]	0.0313	0.0326	0.093	0.35	0.042
t_{\max}	[min]	40.09	43.15	26	1.7	12

		38				
		24 h DDAVP	D-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	930.25	946.19	3100	0.3	1400
k_0	[min ⁻¹]	0.00855	0.0126	0.023	0.56	0.01
t_{\max}	[min]	176.32	104.47	22	4.7	10

		48				
		24 h DDAVP	D-optimal design	95% conf. Interval	95% t-value	standard deviation
D	[U]	50.6	50.57	1200	0.043	530
k_0	[min ⁻¹]	0.0996	0.0988	0.65	0.15	0.29
t_{\max}	[min]	34.54	34.54	9.7	3.6	4.4

As can be read from tables 4.39 and 4.40, the release parameters k_0 , D and t_{\max} can be estimated correctly even from an in silico data set generated with a variance of 36. This result has been graphically expressed in figure 4.58 and 4.59, taking as example subject 48. In figure 4.58, the experimental and predicted profiles of VWF:Ag response generated with a variance of 4 (a) and a variance of 36 (b) following the A-optimal design have been compared. The same comparison has been carried out for D-optimal design and reported in figure 4.59. The predicted profile shows the same value of the peak and the same shape in both pictures 4.58.a and 4.58.b. This is a graphical demonstration that the parameters k_0 , D and t_{\max} have been estimated precisely, even if the error in the experimental data is higher in figure 4.58.b. On the contrary, the position of the peak is different in figure 4.59.a and 4.59.b. Indeed, as readable in table 4.41, the value of D is estimated with uncertainty even if the order of magnitude is correct. However, the difference in the value of the peak between figure 4.59.a and 4.59.b is lower than 20 U/dL, therefore the error is clinically accepted.

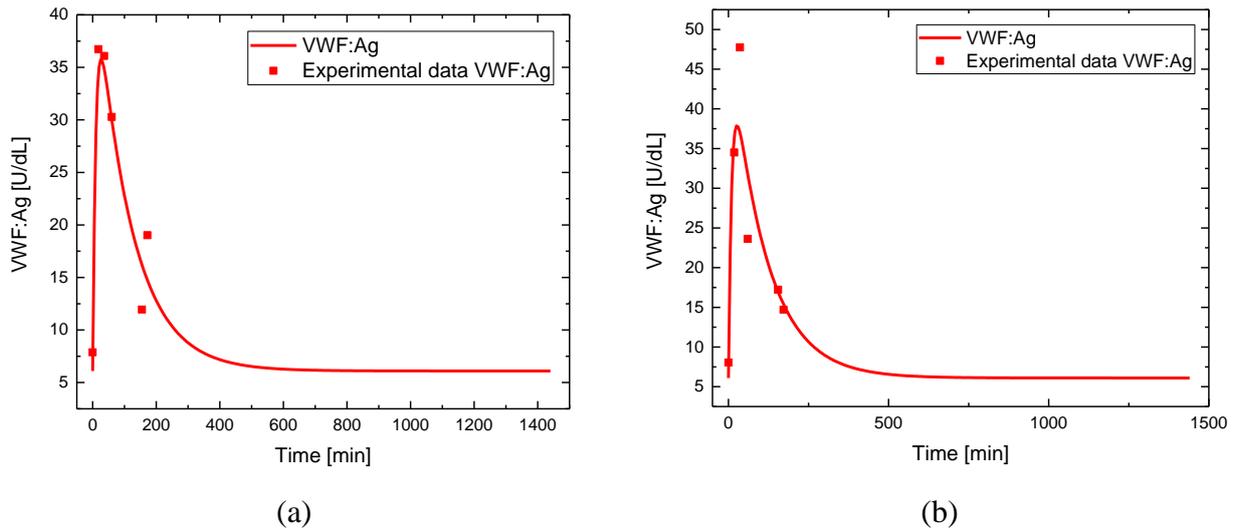


Figure 4.58. Comparison between the experimental and predicted VWF:Ag profiles for patient 48 using the A-optimal design with a variance of 4 (a) and a variance of 36 (b).

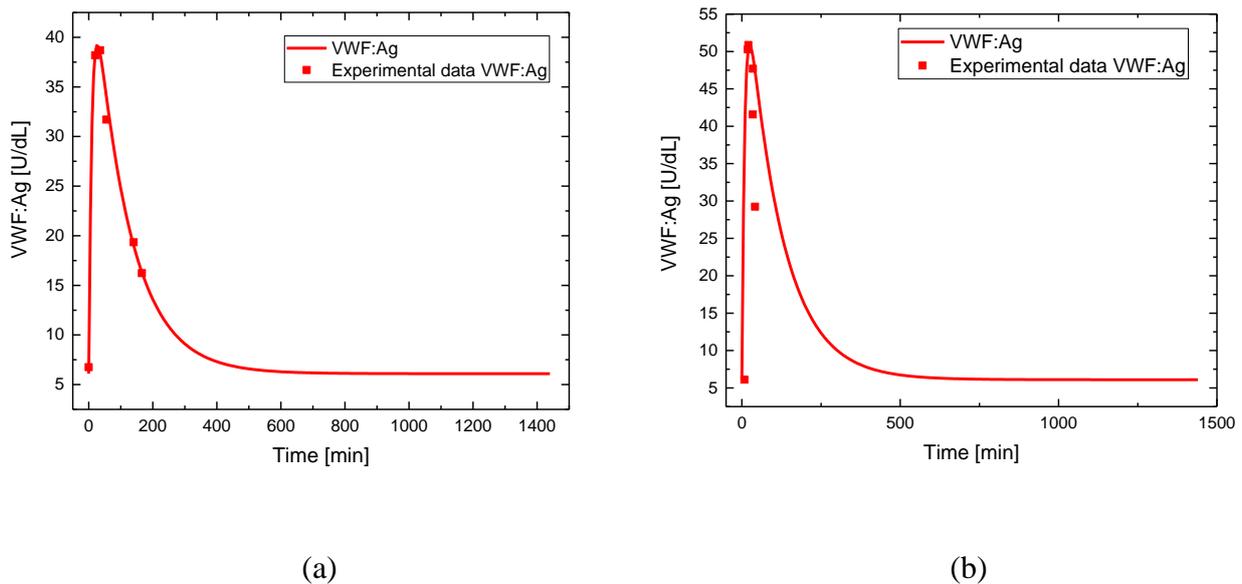


Figure 4.59. Comparison between the experimental and predicted VWF:Ag profiles for patient 48 using the D-optimal design with a variance of 4 (a) and a variance of 36 (b).

The original 24 hours DDAVP protocol has been shortened thanks to the modification of SGM into MGM_2. Indeed, now, only the release parameters k_0 , D and t_{\max} need to be estimated, whereas k_1 and k_e are explicitly calculated from basal quantities. Different time-reduced DDAVP designs have been tested:

- Time-reduced design (simple reduction in time of the original 24 h DDAVP protocol);
- A-optimal design;
- D-optimal design.

All the designs have been tested and validated generating a random set of experimental data for specific patients in each category of VWD and estimating the release parameters k_0 , D and t_{\max} . Then, the robustness of the two optimized designs has also been analysed through the “worst-case” method. As reported in the discussion, all the designs allow to estimate k_0 , D and t_{\max} with precision. Hence, it is impossible to select the best design. Furthermore, as can be clearly visualized in figure 4.60, the three designs are similar in terms of position in time of the sampling points.

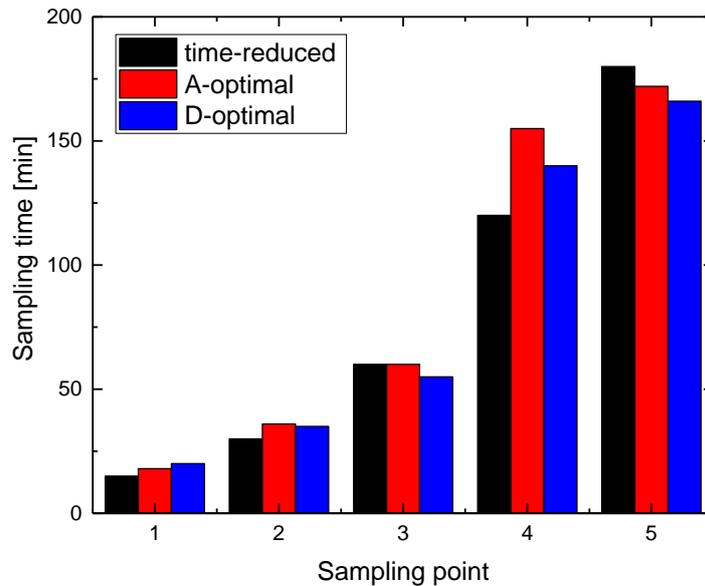


Figure 4.60. Comparison between the position of the samplings in the three considered DDAVP designs: time-reduced, A-optimal and D-optimal design.

The similarity that can be noticed mathematically between the various designs is instead considered equality in the clinical management. In simple terms, for doctors and nurses, the time-reduced design, A-optimal and D-optimal designs are completely identical. In fact, in the everyday clinical procedures, it is impossible to perform measurements with a precision of 10-15 minutes. Therefore, referring to the DDAVP execution, taking samples at 172 minutes (as in A-optimal design) or at 180 min (as in the time-reduced design) is the same. Indeed, an error of ± 8 minutes is totally admitted in the medical procedures. In conclusion, the three tested designs are identical and allow us to get the right estimation of the release parameters. For simplicity, the simple 3 hours time-reduced DDAVP [0, 15, 30, 60, 120, 180] has been used to test MGM_2 in different case studies (§4.2 and §4.4).

4.4 Model representation of out-of-category subjects

Finally, it could be interesting to test the ability of MGM_3_Ag and MGM_2 of representing subjects that do not belong to the already considered categories (HnonO, HO, 2B and Vicenza). The procedure has been tested on patients that belong to other VWD subcategories, which in both cases show a symptomatology (alteration of VWF) like that observed in some of the VWD classes treated in this study. In particular, the subjects that have been considered are four (53, 54, 55 and 56). Two of them (53, 54) are defined as 2B-like patients, meaning that they possess specific mutations, which produce an alteration of VWF similar to that present in normal 2B patients. Instead, subjects 55 and 56 suffer of a particular VWD type 1, which brings the VWF to behave similarly to that of 2B patients.

The four patients have been at first tested with MGM_3_Ag, whose required input data are reported in table 4.41.

Table 4.41. *Required input data for the simulation with out-of-category patients.*

Patients	VWF:Ag	VWF:CB	VWFpp ratio	VWF:R	Weight
53	32.05	29.50	1.89	0.92	60
54	31.50	25.50	1.93	0.81	45
55	29.70	12.90	4.00	0.43	62
56	31.50	16.60	3.25	0.53	65

MGM_3_Ag is able to simulate the subjects only with the basal state correlations for k_0 , k_1 and k_e tailored for 2B category. This is the first important result, because it means that the model MGM_3_Ag can correctly classify the patients. In confirmation of this, it is possible to analyse, in figure 4.61, the relative position of the response surfaces generated for each VWD category in k_0 (a), k_1 (b) and k_e (c) correlations.

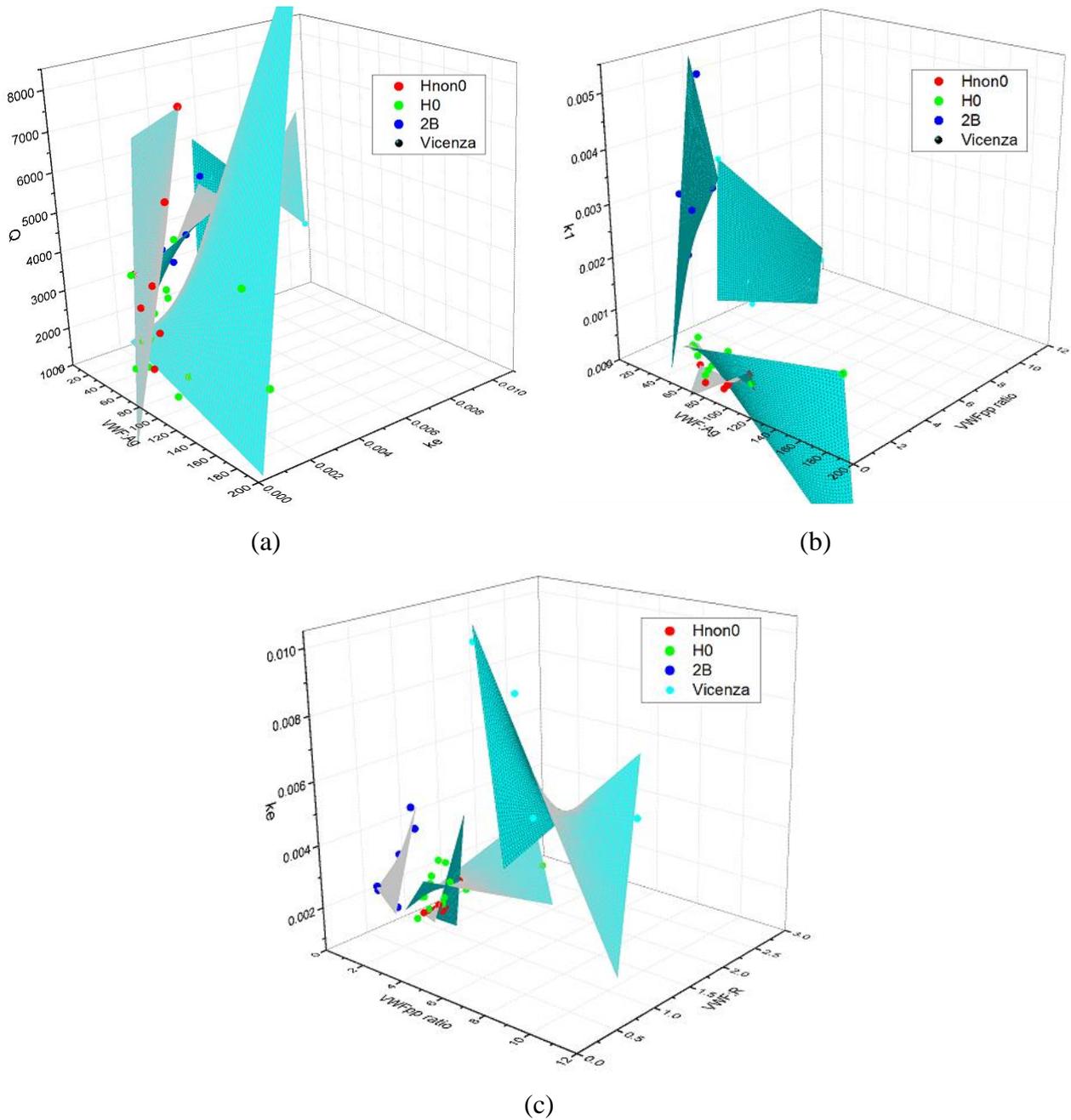


Figure 4.61. Relative position of the response surfaces in k_0 (a), k_1 (b) and k_e (c) correlations.

As is possible to understand, the response surfaces do not cross between each other, meaning that classification with VWD categories can be conducted with MGM_3_Ag.

However, as illustrated in figure 4.62, MGM_3_Ag is not able to reproduce correctly the VWF:Ag and VWF:CB profiles, because the error between the experimental and simulated trends is relevant for all the patients.

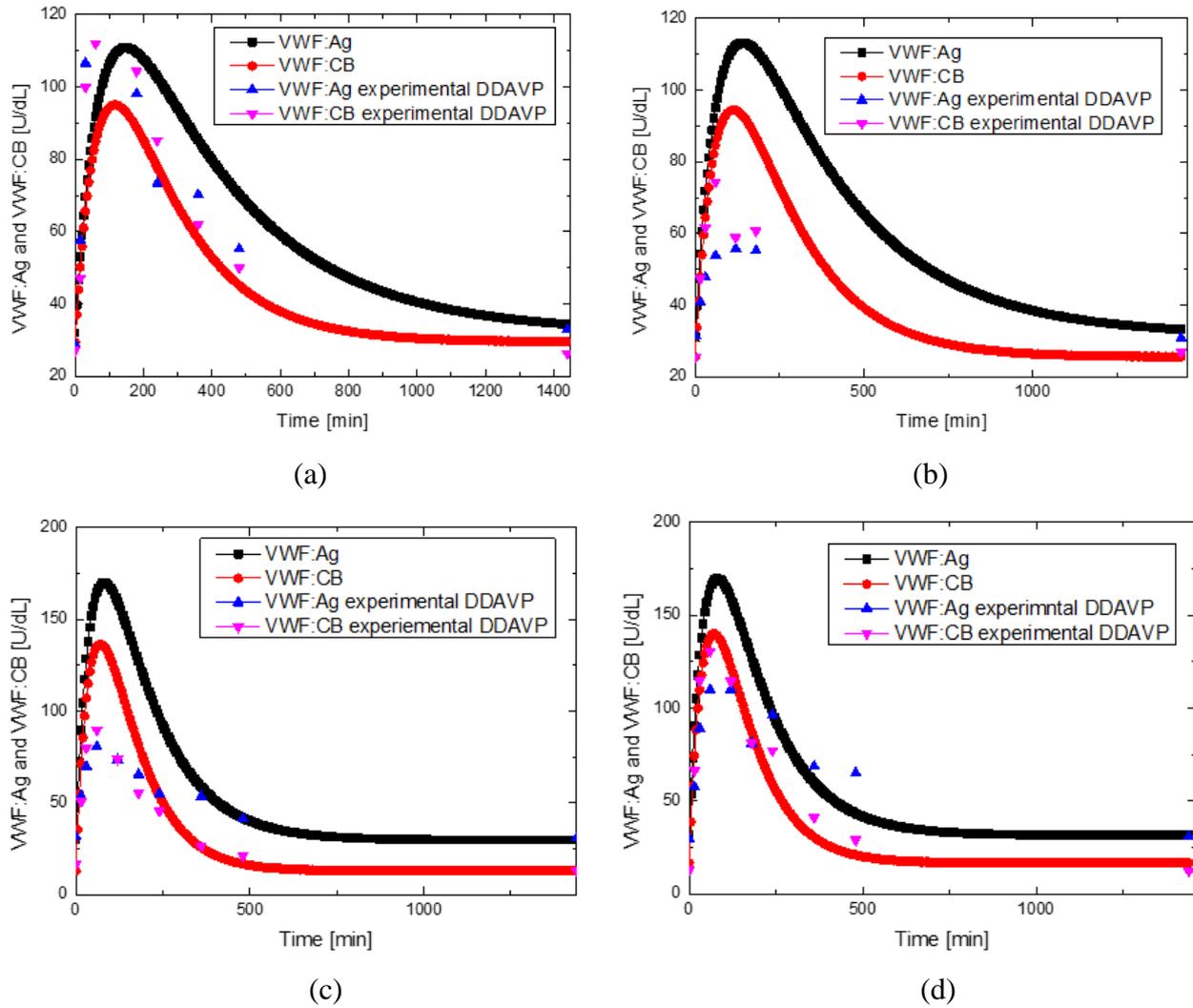


Figure 4.62. Comparison between experimental and simulated VWF:Ag and VWF:CB profiles with MGM_3_Ag for patients 53 (a), 54 (b), 55 (c) and 56 (d).

The error is higher in the release path and in particular in the position of the peak. This observation reinforces the already known deduction that D and t_{\max} cannot be used at the average value of each category. Therefore, the approach defined in the thesis has been used. The time-reduced DDAVP [0, 15, 30, 60, 120, 180] has been applied to estimate the release parameters k_0 , D and t_{\max} for patients 53, 54, 55 and 56. Then, the estimated k_0 , D and t_{\max} have been substituted in MGM_2 and the VWF:Ag and VWF:CB have been simulated and compared with the experimental DDAVP data. Results are reported in figure 4.63 and as can be clearly visualized, MGM_2 allows to reproduce the profiles with a lower error.

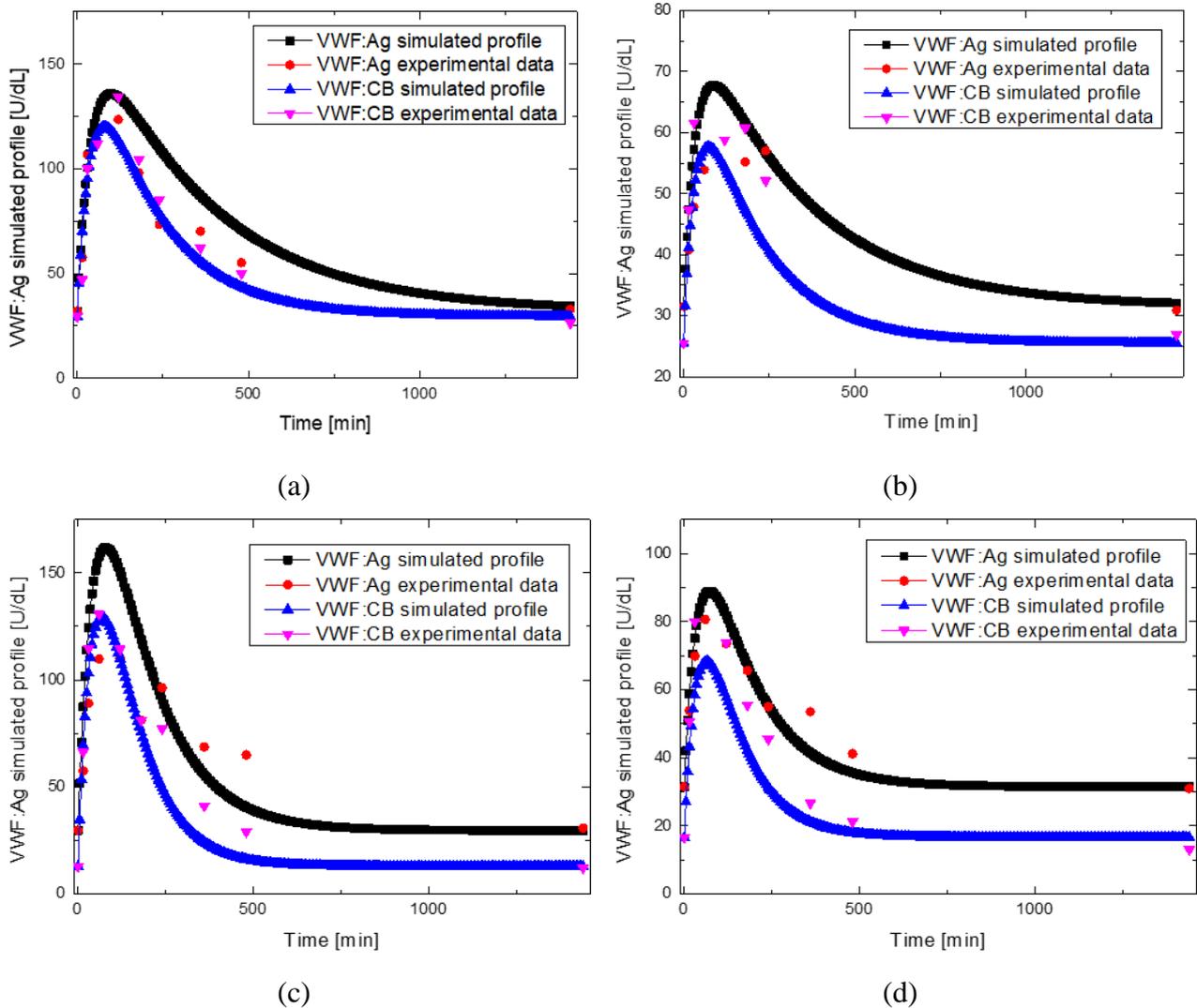


Figure 4.63. Comparison between experimental and simulated VWF:Ag and VWF:CB profiles with MGM_2 for patients 53 (a), 54 (b), 55 (c) and 56 (d).

Even in this case the profiles cannot perfectly overlap with the experimental data, because the patients do not strictly belong to the four categories considered in this study.

Of course, the target of reproducing subjects out-of-category with MGM_2 is not to make diagnosis, but to observe, once the diagnosis is known, some biochemical details, that can help physicians in the characterization of VWF path. In order to make a reliable model-based diagnosis and characterization for these particular subjects, MGM_2 should be extended also to VWD type 1 and to other VWD classes (2N, 2M, 2A I, 2A II) and subcategories. Future work will be indeed oriented on this extension in order to make the model more general.

Conclusions

The diagnosis of VWD is a complex task made more difficult by the various number of VWD subtypes (type 1, 2A, 2B, Vicenza and others). PK models have been recently proposed for disease classification and characterization. The big advantage of using PK models to help in the diagnostic process of VWD is that it is possible to quantify the mechanisms of:

- VWF release (k_0)
- VWF proteolysis (k_1)
- VWF elimination (k_e)

At first, SGM has been proposed as PK model for disease characterization and classification. The SGM works well for the estimation of the PK parameters, but a 24 hours DDAVP needs to be carried out to achieve a statistically satisfactory estimation of the disease metabolic parameters. However, the main issue is that a 24 hours DDAVP test is required for model identification and this is a long and invasive non-routine test. Therefore, in this Thesis a way for achieving model identification only from basal clinical trials has been studied. The alternative basal tests considered in this study are:

- VWFpp, to quantify VWF elimination from the bloodstream
- VWF:Ag to evaluate the number of VWF antigens in the bloodstream
- VWF:CB to analyse the VWF in binding with collagen

From them, two other physiological quantities are derived:

- VWFpp ratio, expressed as $VWFpp/VWF:Ag$
- VWF:R, defined as $VWF:CB/VWF:Ag$

Response surface methodology has been applied for the development of suitable correlations, which relate explicitly the kinetic parameters k_0, k_1, k_e of model SGM with basal clinical trials.

- $k_0 \rightarrow f(VWF:Ag, VWF:R, \text{intraplatelet VWF})$
- $k_1 \rightarrow f(VWF:Ag, VWFpp \text{ ratio})$
- $k_e \rightarrow f(VWFpp \text{ ratio}, VWF:R)$

Then, the new equations have been substituted into SGM. In particular, three levels of model modification have been proposed:

- MGM_1, in which only the correlation for the elimination kinetic constant k_e has been added in the equation set of SGM;
- MGM_2, in which correlations for k_e and k_1 have been inserted in SGM;
- MGM_3, where all the kinetic parameters k_0, k_1, k_e are calculated directly from basal data.

The three models have been tested and results demonstrated that MGM_1 and MGM_2 work well for disease characterization, whereas MGM_3 produces a sensible error in the release description. Indeed, the comparison between MGM_1, MGM_2 and SGM shows that the profiles overlap. On the contrary, the error produced by the simulation of VWF:Ag and VWF:CB responses with MGM_3 is high for some subjects if compared with the profiles generated by SGM. The reason can be found in the assumption made for model modification. Precisely, D and t_{\max} , which are two other relevant quantities for describing the release, cannot be used at the average value of each category, but they must be specific for each subject. The problem is that explicit correlations for D and t_{\max} cannot be defined from basal trials. Hence, it is impossible to describe the release precisely with MGM_3. However, the results obtained from model modification highlighted the possibility to redesign the DDAVP based on MGM_2 and to sensibly reduce the test execution time.

In conclusion, the most ambitious target of the project, that is the elimination of the DDAVP execution for model identification has not been achieved completely. Indeed, even if suitable correlations for the calculation of k_0 at the basal state have been defined, the uncertainty on the average values of the parameters D and t_{\max} does not allow to reproduce VWF:Ag and VWF:CB profiles with precision as in SGM. However, important achievements have been reached throughout this work: the SGM model has been modified into MGM_2, that explicitly evaluates k_1 and k_e from basal quantities; the release parameters still need to be estimated through the DDAVP clinical trial, but the DDAVP duration has been successfully reduced to three hours. The reduction in time of the clinical trial is remarkable because it strongly improves the quality of life of the patients that undergo a less stressful clinical procedure and it facilitates the clinical management considering both economical and organizational aspects.

Future work will be carried out to define clearly the regions of validity of the model among the different categories and to reinforce model validation by applying MGM_2 and the time-reduced DDAVP to characterize the VWF:Ag and VWF:CB profiles of new patients, which do not belong to the pool of subjects already considered. Furthermore, the results obtained through this thesis could represent the first step towards the implementation of a software that will be developed to help medical doctors in the VWD diagnosis and characterization.

Symbols

A	first RSM parameter	$[\text{min}^{-1}]$
B	second RSM parameter	$[\text{min}^{-1}]$
C	third RSM parameter	$[\text{min}^{-1}]$
\underline{D}	fourth RSM parameter	$[\text{min}^{-1}]$
k_0	kinetic constant of release	$[\text{min}^{-1}]$
k_1	kinetic constant of proteolysis	$[\text{min}^{-1}]$
k_e	kinetic constant of elimination	$[\text{min}^{-1}]$
D	amount of VWF released	$[U]$
$x^{\text{UL+HMW}}$	ultra large and high molecular weight multimers units	$[U]$
$x_b^{\text{UL+HMW}}$	ultra large and high molecular weight multimers units (basal state)	$[U]$
x^{LMW}	low molecular weight multimers units	$[U]$
x_b^{LMW}	low molecular weight multimers units	$[U]$
y^{Ag}	antigen concentration	$[U/dL]$
y^{CB}	collagen binding concentration	$[U/dL]$
y_b^{Ag}	antigen concentration at the basal state	$[U/dL]$
y_b^{CB}	collagen binding concentration at the basal state	$[U/dL]$
V_d	approximate distribution volume	$[\text{mL}/\text{kg}_{\text{BW}}]$
k	correction factor	$[-]$
y_{CB}'	corrected collagen binding concentration	$[U/dL]$

Greek letters

σ^2	variance	$[-]$
------------	----------	-------

Vectors and matrices

\mathbf{q}	sensitivity vector	$[N_\theta]$
\mathbf{H}_θ	Fisher information matrix	$[N_\theta \times N_\theta]$
θ^0	initial parameter set	$[N_\theta]$

θ	parameter set	$[N_\theta]$
θ'	perturbed parameter set	$[N_\theta]$
q_i^{Ag}	sensitivity on antigen concentration	$i = 1, \dots, N_\theta$
q_i^{CB}	sensitivity on collagen binding concentration	$i = 1, \dots, N_\theta$
$tr[\mathbf{H}_\theta]$	trace of FIM	$[-]$

Acronyms

VWD	von Willebrand disease
VWF	von Willebrand factor
VWF:Ag	VWF antigen concentration
VWF:CB	VWF collagen binding concentration
VWF:RCo	VWF ristocetin cofactor activity
VWFpp	VWF propeptide
VWFpp ratio	VWF propeptide ratio
VWF:R	VWF collagen binding and antigen concentration ratio
RIPA	ristocetin induced platelet adhesion
FVIII	factor eight
DDAVP	desmovasopressin
UL+HMW	ultra large plus high molecular weight
LMW	low molecular weight
RSM	response surface methodology
PK	pharmacokinetic
FIM	Fisher information matrix
SSE	sum of squared errors
SST	total sum of squares
SSR	residual sum of squares
RMSE	root mean squared errors
PSE	Process systems enterprise
DAE	Differential and algebraic equation

Acknowledgments

Dr. Federico Galvanin

per il suo support, la sua disponibilità costante e per avermi fatto conoscere e apprezzare gli strumenti che hanno reso possibile il lavoro di tesi.

Prof. Fabrizio Bezzo

per credere in me, per avermi dato e darmi in futuro importanti opportunità.

Prof. Alessandra Casonato

per avermi fornito i dati sperimentali che hanno portato al raggiungimento dei risultati presentati nella tesi.

Mamma

per avermi asciugato lacrime e condiviso gioie in questi anni, per essere la mia certezza sempre, perché oggi ti laurei veramente anche tu con me.

Papà

per avermi insegnato il sacrificio e la dedizione al lavoro, per credere sempre in me e nelle mie scelte, per avermi accompagnato ad ogni esame, in modo tale che fosse un successo.

Amici, pochi ma buoni

Per apprezzare i miei pregi e tollerare i miei difetti, per la stima e l'affetto che da anni mi fate sentire.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*(19), 716-723.
- Bard, Y. (1974). *Nonlinear parameter estimation*. Academic Press.
- Berntrop, E. (2007). Erik von Willbrand. *Thrombosis research* (120), S3-S4.
- Box, G., & Behnken, D. (1960). Some new 3 level designs for the study of quantitative variables. *Technometrics*(2), 455-475.
- Box, G., & Lucas, H. (1959). Design of experiments in non-linear situations. *Biometrika*(46), 77-90.
- Casonato, A., Pontara E., Sartorello F. (2006). Identifying type Vicenza von Willebrand disease. *J Lab Clinical Medicine*(147), 96-102.
- Casonato, A., Daidone, V., Padrini R. (2011). Assesment of von Willebrand factor propeptide improves the diagnosis of von Willebrand disease. *Seminary Thrombosis Haemostasis*(37), 456-463.
- Espie, D., & Macchietto, S. (1989). The optimal design of dynamic experiments. *AICHE Journal*(35), 223-229.
- Fedorov, V., Leonov, S. (2014). *Optimal design for nonlinear response models*. Boca Raton: Taylor & Francis Group.
- Galvanin, F., Ballan, C., Barolo, M., Bezzo, F. (2013). A general model-based design of experiments approach to achieve practical identifiability of pharmacokinetic and pharmacodynamic models. *Journal of Pharmacokinetic and Pharmacodynamic*(40), 451-467.
- Galvanin, F., Barolo, M., Padrini, R., Casonato, A., Bezzo, F. (2014). A model-based approach to the automatic diagnosis of von Willebrand disease. *AICHE Journal*(60), 1718-1727.
- Groot, E., Fijnheer, R., Sebastian, S. (2009). The active conformation of von Willebrand factor in patients with thrombotic thrombocytopenic purpura in remission. *Journal of Thrombosis and Haemostasis*(7), 962-969.
- Lillicrap, D. (2007). Von Willebrand disease-phenotype versus genotype: Deficiency versus disease. *Thrombosis Research*(87), 57-64.

- Miao, H., Xia, X., Perelson, A., Wu, H. (2011). On identifiability on nonlinear ODE models and applications in viral dynamics. *SIAM Review*(53), 3-39.
- Montgomery, D. (2001). *Design and analysis of Experiments*. Richmond, Texas, USA: Wiley.
- OrigineLab Corporation. (2007). *Origine User Guide 8*. Northampton, USA.
- Process Systems Enterprise. (2004). *gPROMS advanced user guide*. London: Process Systems Enterprise Ltd.
- Process Systems Enterprise. (2004). *gPROMS introductory user guide*. London: Process Systems Enterprise Ltd.
- Sadler, J. E. (2003). Von Willebrand disease type 1: a diagnosis in search of a disease. *Blood*(101), 2089-2093.
- Sweeney, J. D., Hoerning, L.A. (1992). intraplatelet von Willebrand factor and AB0 blood group. *Thrombosis Research*(68), 393-398.
- Zaverio, M. (2007). The role of von Willebrand factor in thrombus formation. *Thrombosis Research*(120), S9-S16.
- Zieger, B., Budde, U., Jessat, U., Zimmermann, R., Simon, M., Katzel, R., Sutor, A.H. (1997). New Families with von Willebrand disease type 2M (Vicenza). *Thrombosis Research*(87), 57-64.

Appendix A

A.1 Available dataset for healthy subjects

Table A1. Punctual data for Hnon0 and H0 categories.

Patient	VWF:Ag	VWF:CB	VWFpp ratio	VWF:R	k_e	k_0	k_1	Diagnosis
1	108.9	119.3	0.714	1.095	3.453E-04	5.74E-04	0.025	HnonO
2	80.6	101.5	0.962	1.259	6.828E-04	1.69E-07	0.024	HnonO
3	86.80	91.90	0.768	1.059	8.057E-04	1.61E-04	0.023	HnonO
4	85.00	93.60	0.953	1.101	7.199E-04	1.46E-04	0.038	HnonO
5	122.20	131.70	0.696	1.077	4.200E-04	7.12E-05	0.029	HnonO
6	58.3	88.4	1.288	1.516	5.716E-04	4.41E-04	0.055	HnonO
7	51.70	57.20	1.41	1.106	1.177E-03	1.19E-04	0.028	HnonO
9	87.7	98.4	0.916	1.122	4.825E-04	5.70E-04	0.025	HnonO
10	79.90	83.70	1.047	1.047	5.212E-04	1.69E-07	0.024	HnonO
11	104.50	155.20	1.028	1.485	1.510E-03	4.45E-04	0.028	HnonO
12	105.80	119.20	1.125	1.126	9.355E-04	2.53E-04	0.023	HnonO
13	217.30	311.00	1.345	1.431	1.151E-04	3.09E-04	0.026	HnonO
14	105.70	139.10	0.876	1.316	7.293E-04	5.56E-04	0.034	HnonO
15	65.80	68.50	0.855	1.041	8.531E-04	1.72E-07	0.026	HnonO
16	99.20	110.00	0.858	1.108	9.0041E-04	2.10E-07	0.0172	HO
17	107.20	110.00	0.828	1.026	1.083E-03	3.84E-04	0.016	HO
18	56.20	67.80	1.651	1.206	1.896E-03	2.24E-07	0.035	HO
19	58.10	56.50	1.196	0.972	3.037E-03	1.77E-03	0.044	HO
20	54.50	145.20	1.147	2.664	8.187E-04	3.19E-03	0.031	HO
21	62.2	72.1	1.397	1.160	1.146E-03	2.18E-07	0.041	HO
22	53.20	69.20	1.073	1.300	2.337E-03	5.33E-04	0.027	HO
23	79.50	119.05	1.032	1.497	1.131E-03	5.61E-05	0.029	HO
24	61.00	66.70	1.133	1.093	2.109E-03	4.99E-05	0.024	HO
25	86.00	104.70	0.910	1.217	1.587E-03	7.77E-04	0.021	HO
26	80.6	110	0.911	1.364	1.633E-03	2.22E-07	0.020	HO
27	183.80	204.00	0.609	1.109	1.262E-03	1.33E-03	0.023	HO
28	55.60	74.40	1.081	1.338	9.170E-04	3.96E-04	0.019	HO
29	50.60	54.30	1.130	1.073	1.871E-03	5.46E-04	0.018	HO
30	159.20	164.70	0.799	1.034	1.388E-03	2.25E-07	0.015	HO
31	58.70	75.20	1.332	1.281	1.205E-03	9.92E-04	0.021	HO

32	63.20	93.80	1.389	1.484	1.285E-03	2.18E-04	0.046	HO
33	58.80	68.20	1.263	1.159	2.602E-03	2.10E-03	0.025	HO
34	62.10	53.8	1.241	0.866	9.465E-04	1.44E-04	0.020	HO
35	53.00	64.30	1.279	1.213	1.277E-03	6.96E-04	0.034	HO

A.2 Available dataset for 2B and Vicenza subjects

Table A2. Punctual data for 2B and Vicenza categories.

Patient	VWF:Ag	VWF:CB	VWFpp ratio	VWF:R	k_e	k_0	k_1	Diagnosis
36	17.50	9.25	2.77	0.528	4.573E-03	0.021	0.02015	2B
37	39.60	9.30	1.98	0.234	2.763E-03	0.003	0.01970	2B
38	43.00	20.7	2.1	0.481	3.673E-03	2.910E-06	0.00855	2B
39	34.60	12	1.65	0.346	2.182E-03	0.00321	0.02617	2B
40	42.90	10.6	2.9	0.247	2.405E-03	0.00321	0.02239	2B
41	19.89	9.7	2.79	0.487	5.279E-03	0.00179	1.61E-02	2B
42	48.00	10.6	2.00	0.220	2.903E-03	0.00537	0.02227	2B
43	11.25	8.55	5.05	0.760	1.033E-02	0.00330	0.06932	Vicenza
44	18.10	18.5	11.33	1.022	8.633E-03	7.149E-05	0.036998	Vicenza
45	6.90	5.6	6.7	0.811	9.037E-03	0.00059	0.06059	Vicenza
46	10.22	10.40	11.2	1.017	6.044E-03	0.00040	0.09123	Vicenza
47	7.99	8.45	10.42	1.057	8.346E-03	6.992E-05	0.10348	Vicenza
53	8.69	8.5	7.20	0.978	9.323E-03	5.528E-05	0.10481	Vicenza
48	6.10	5.49	7.08	0.900	5.375E-03	2.369E-05	0.09960	Vicenza
49	7.20	3.60	10.85	0.500	9.736E-03	0.00895	0.01920	Vicenza

Appendix B

B.1 Investigation of the response surface for parameter k_e

Thanks to the collaboration with the medical school, it has been suggested to investigate a possible correlation between the elimination kinetic parameter k_e and VWFpp ratio. Indeed, commonly blood coagulation experts exploit VWFpp ratio quantity to indirectly measure the elimination of VWF from the blood stream. Analysing the relation between VWFpp ratio and k_e (fig. B.1), clusters are clearly visible between the different categories.

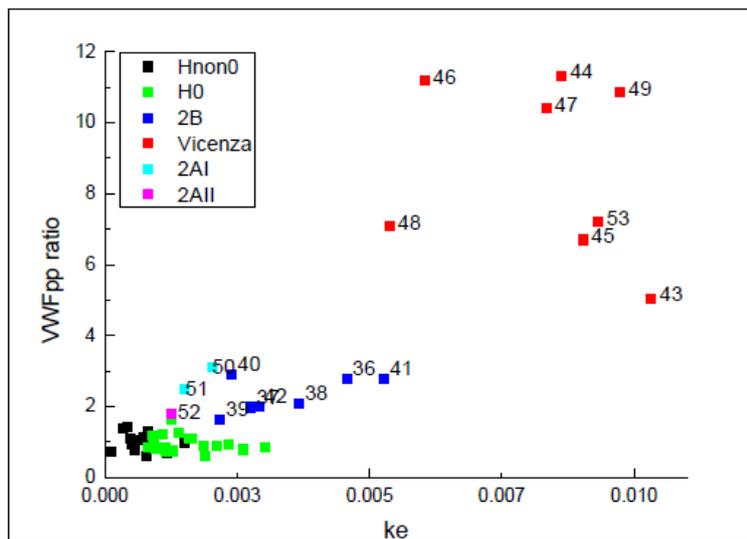


Figure B.1. Correlation between VWFpp ratio and k_e .

As expected, Vicenza subjects show a higher VWFpp ratio compared to the other considered categories. This finding is due to the low VWF:Ag level in the blood stream usual in Vicenza category, which is caused by an increased elimination constant k_e . Furthermore, as illustrated in figure B.2 also the correlation between k_e and VWF:R has been analysed. The picture clearly shows that healthy subjects have a VWF:R value around 1, whereas unhealthy subjects slightly lower with a higher elimination constant.

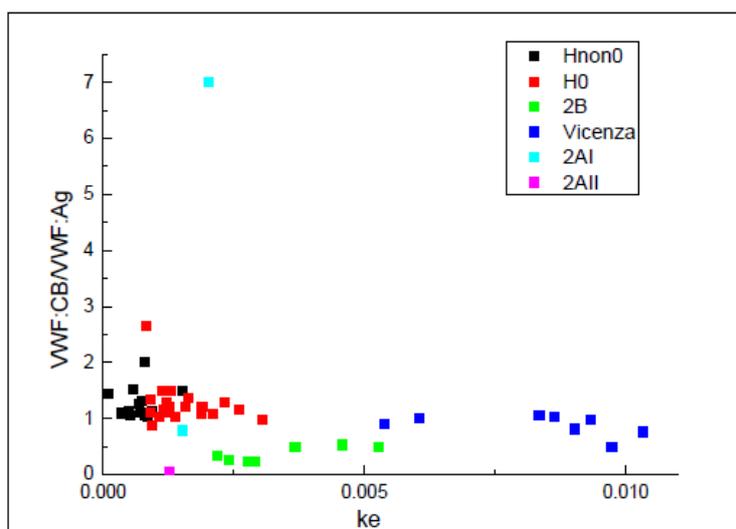


Figure B.2. Correlation between VWF:R and k_e .

Looking at figure B.1, linear trends between the categories might be considered. However, as figures B.3, B.4, B.5 and B.6 demonstrate a one factor linear correlation for k_e does not work.

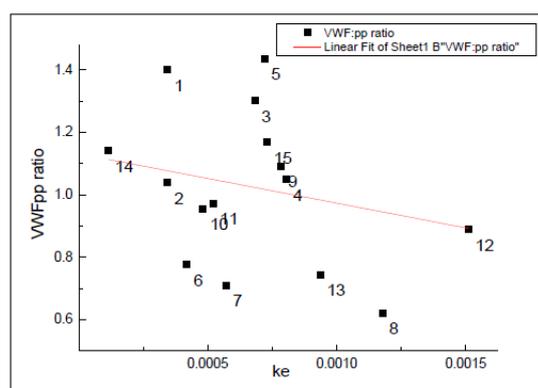


Figure B.3. Linear fitting of VWF:pp ratio and k_e considering Hnon0 category.

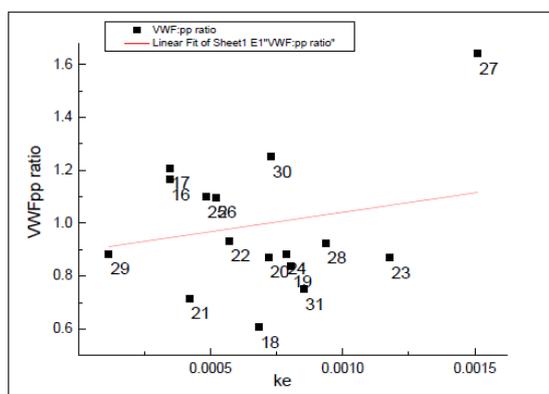


Figure B.4. Linear fitting of VWF:pp ratio and k_e considering H0 category.

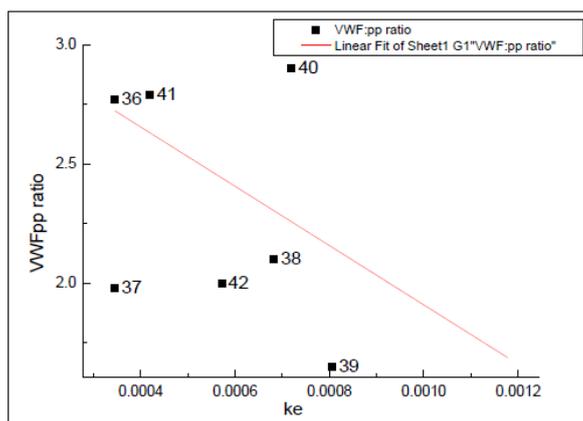


Figure B.5. Linear fitting of VWF:pp ratio and k_e considering 2B category.

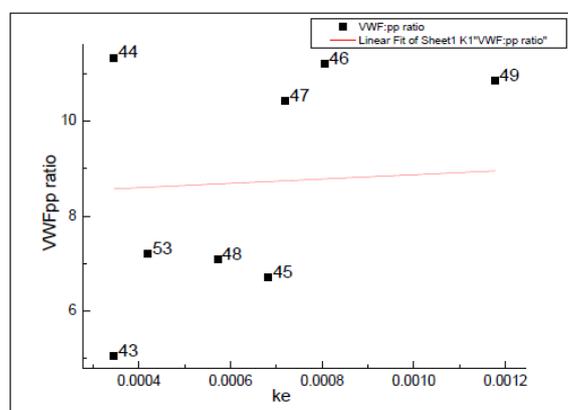


Figure B.6. Linear fitting of VWF:pp ratio and k_e considering Vicenza category.

As is possible to understand from the results obtained a one factor correlation is not sufficient to allow the calculation of the elimination constant from basal state clinical quantities. Therefore, RSM has been applied for studying a suitable correlation for k_e as function of VWF:pp ratio and VWF: R. In particular, two analytical forms of the correlation have been investigated. As seen in §4, linear response surface with interactions has been chosen to modify the SGM (2017) into MGM_1. However, also the quadratic response surface with interactions has been studied as possible mathematical form of the correlation.

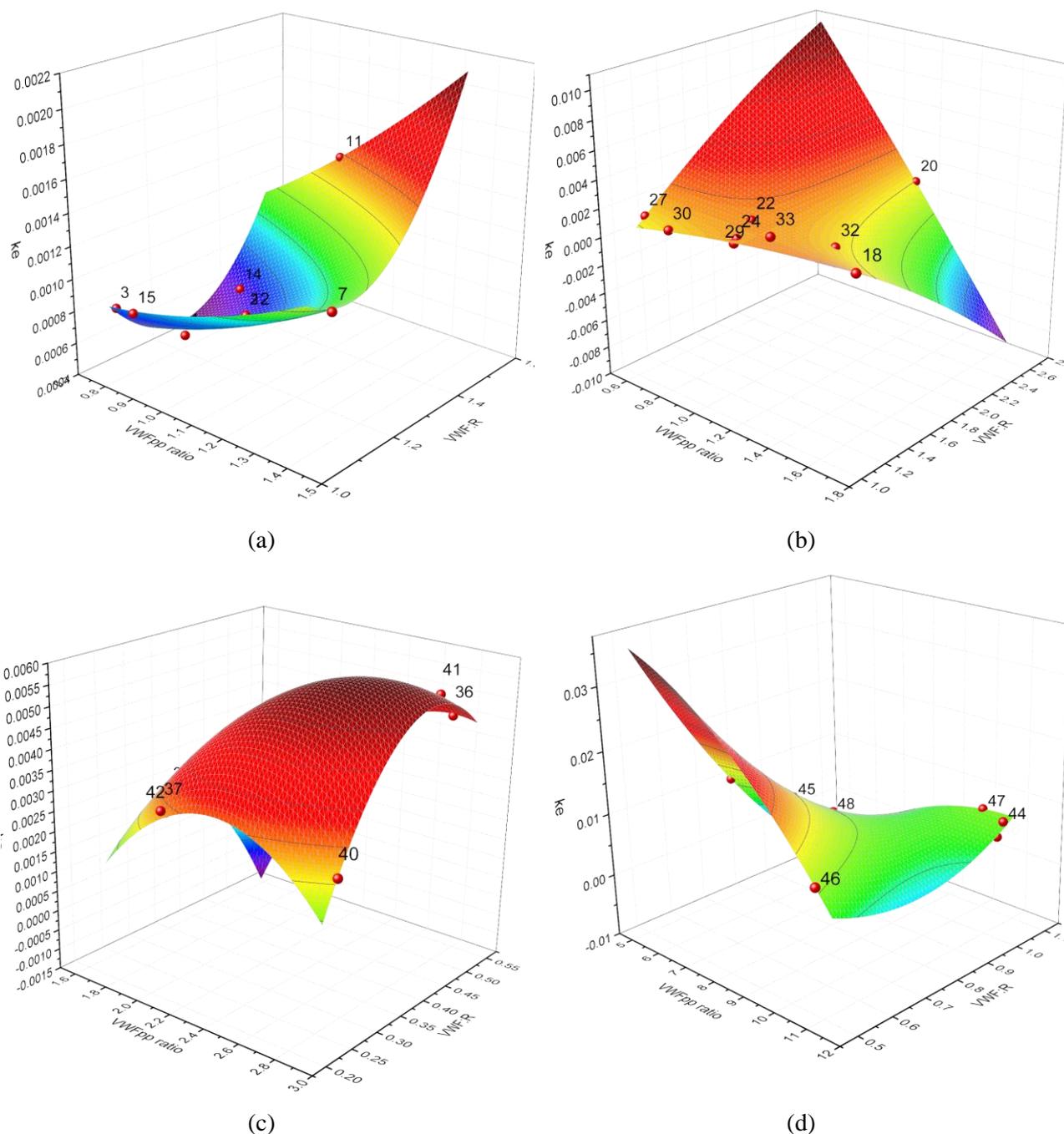


Figure B.7. Quadratic fitting for parameter k_e HnonO (a), HO (b), 2B (c) and Vicenza (d) categories.

However, even if the fitting appears to work well (from both visual and statistical analysis) for all the categories of disease, the choice goes on the linear response surface with interaction instead of on the quadratic response surface, because the AIC is lower (see table B.1). The AIC is a good index for the evaluation of the quality of fitting (§2.2.4), and its definition states that the best model is the model with the lowest number of parameters but able to produce a high-quality fitting.

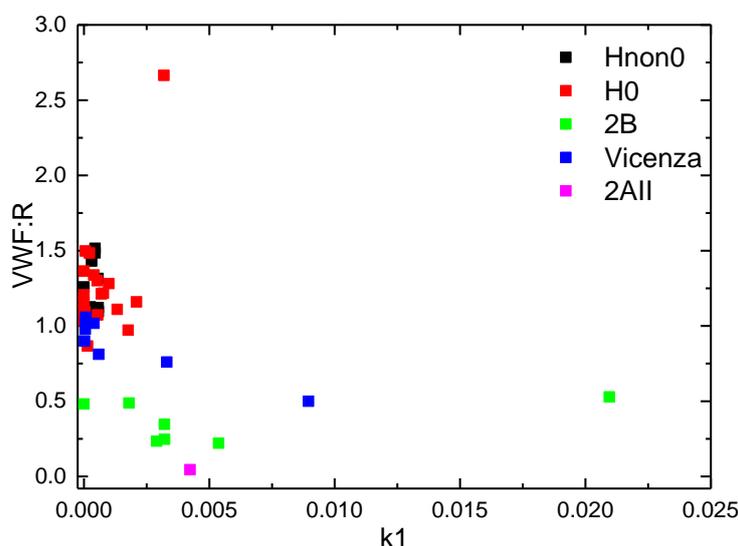
Table B.1. AIC index for linear and quadratic response surface with interactions.

	AIC	
	Linear response surface	Quadratic response surface
HnonO	4.64E+01	4.74E+01
HO	4.10E+01	4.44E+01
2B	4.06E+01	4.21E+01
Vicenza	3.33E+01	4.44E+01

In conclusion, visual, mathematical and statistical analysis confirms that linear response surface with interactions is the best structure for model modification of the system considered in the study.

B.2 Investigation of the response surface for parameter k_1

Several basal clinical quantities can be meaningfully related to the definition of the correlation for the proteolytic parameter k_1 . In particular, some of the clinical measurements that can be related to the proteolytic parameter are: VWFpp ratio, VWF:R and VWF:Rco. As figure B.8 illustrates, the analysis of the one factor correlation between VWFpp ratio and k_1 shows clusters between the different categories but trends are clearly not recognizable. This means that, as for k_e , a one factor correlation is not sufficient for the calculation of k_1 using basal state quantities. Therefore, RSM needs to be applied.

**Figure B.8.** Correlation between VWF:R and k_1 .

Considering that both k_1 and k_e are related to VWF:R and VWFpp ratio, a linear response surface with interactions has been developed for k_1 as function of k_e and VWF:R. However, as is visible from

figure B.9 and from the R^2 in table B.2, the results of the fitting are completely not satisfying, meaning that other correlations need to be investigated.

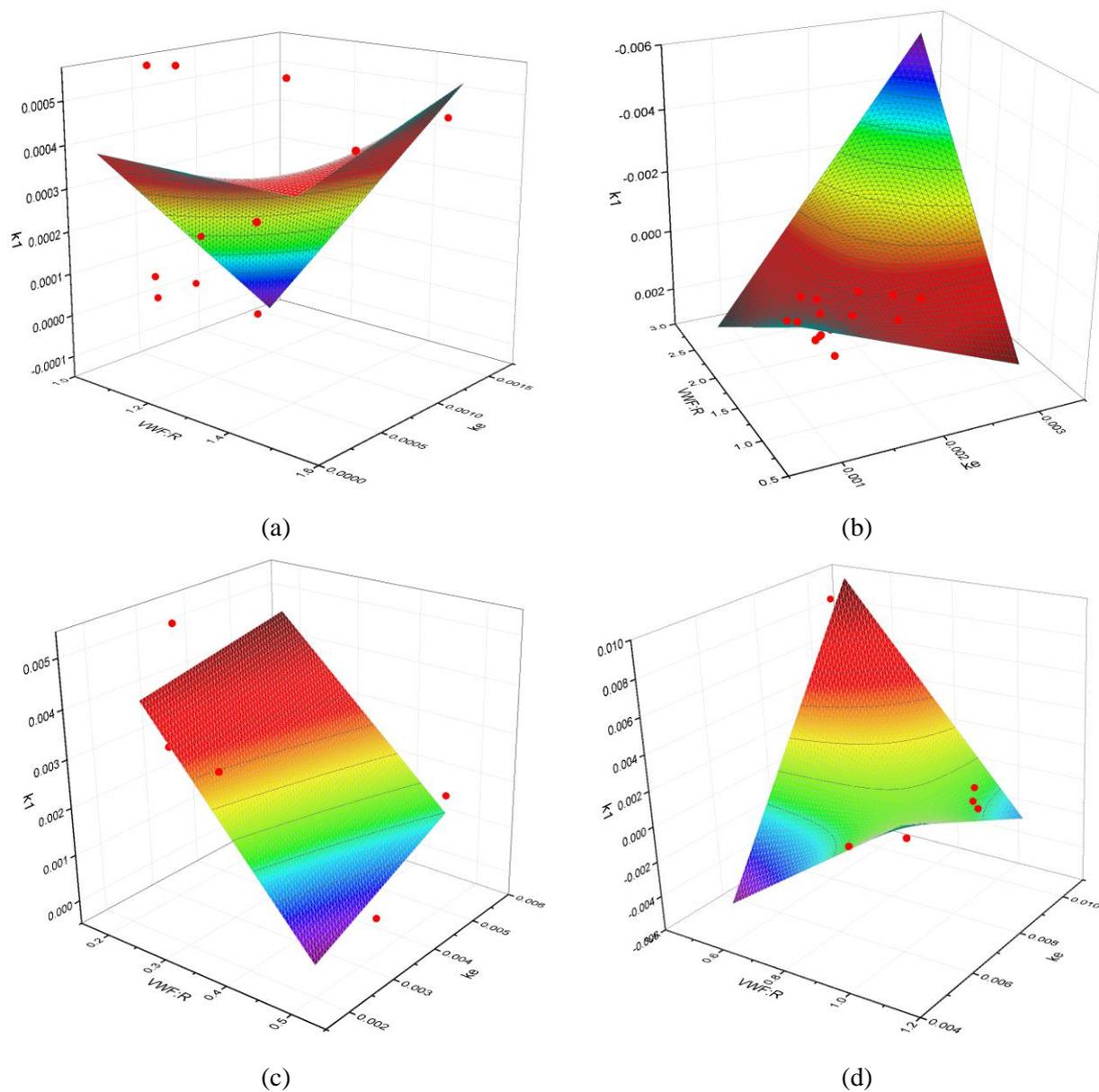


Figure B.9. Linear response surface with interactions considering HnonO (a), HO (b), 2B (c) and Vicenza (d) categories.

Table B.2. R^2 values for the different categories.

	R^2
HnonO	0.56164
HO	0.79905
2B	0.58816
Vicenza	0.96522

It is known that VWF:Rco is one of the most important assays for the evaluation of the proteolytic behaviour of VWF. Therefore, RSM has been applied finding a new explicit correlation for k_1 as function of VWF:R and VWF:Rco. The form of the correlation is again a linear response surface with interactions. The correlation has been tested for HO, 2B and Vicenza categories because VWF:Rco experimental values for HnonO category were not available. As visible from figure B.10 and table B.3, results of the fitting are really satisfying.

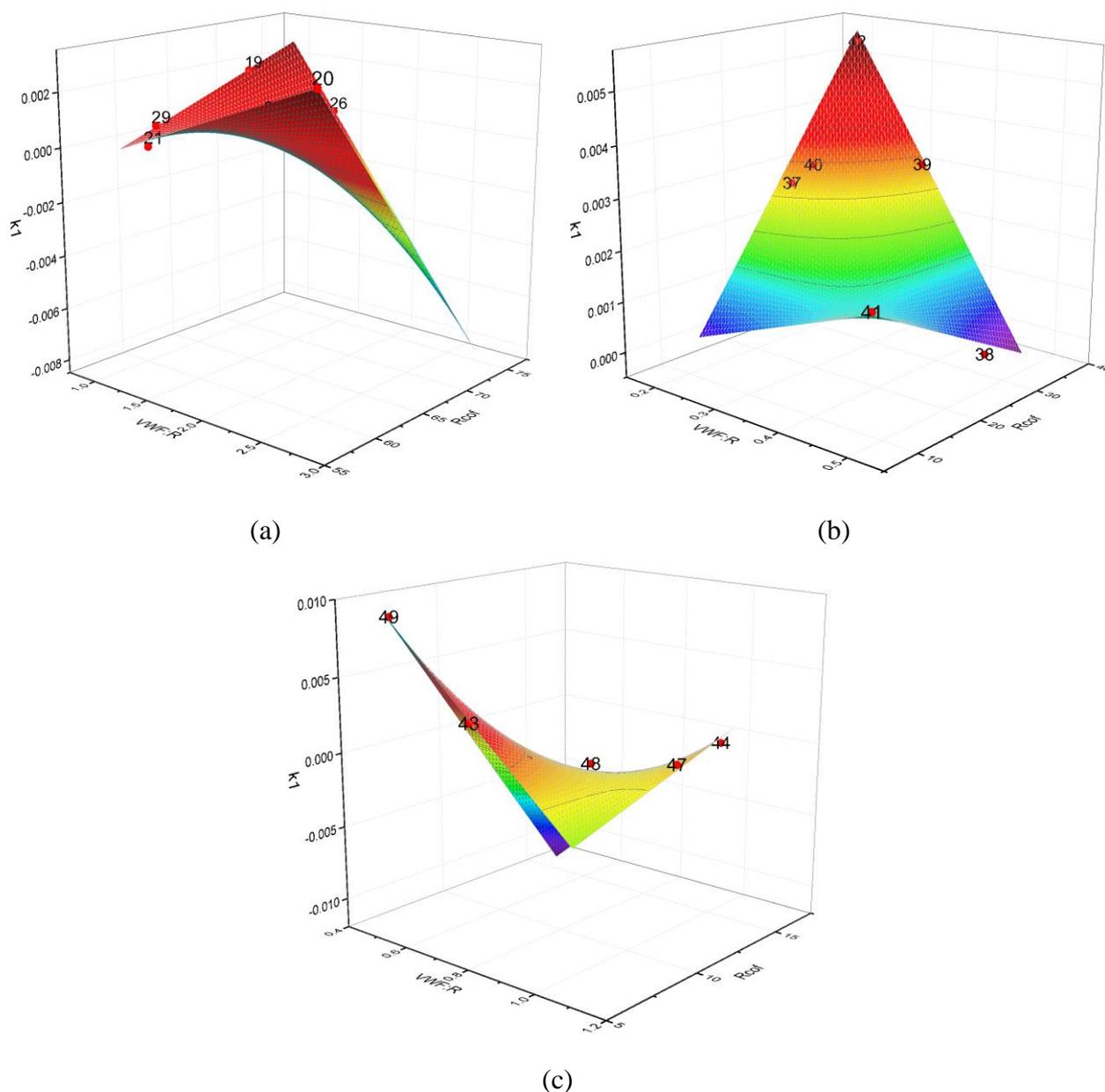


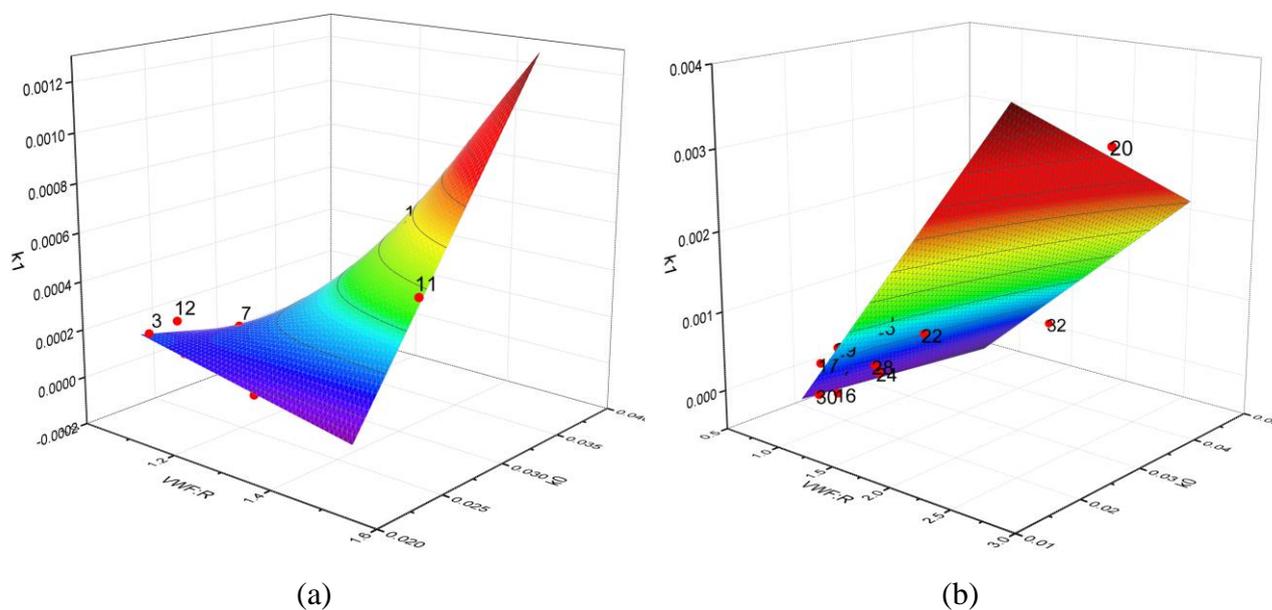
Figure B.10. VWF:Rco linear response surface with interactions considering HO (a), 2B (b) and Vicenza (c) categories.

Table B.3. R^2 values for the different categories in VWF:Rco correlation.

	R^2
HO	0.99132
2B	0.98601
Vicenza	0.98931

However, the target of the project is not only the reduction of the time required by the DDAVP execution, but also to simplify the diagnosis and characterization of VWD. Therefore, simplification implies not only the reduction of the DDAVP execution time, but also the definition of possible correlations, which depend on clinical trials easy to conduct. To respect this purpose, VWF:Rco must not be adopted in the definition of k_1 . Indeed, VWF:Rco is a time-demanding assay that requires specific instrumentations and a good ability in platelets management to limit the errors.

Starting from the SGM a mass balance on a restricted control volume (UL+HMW multimers) has been executed. This allows us to find a new possible correlation for k_1 as function of VWF:R and k_0 . As figure B.11 illustrates and as it can be read from table B.4, the fitting and the related statistics appear to be good.



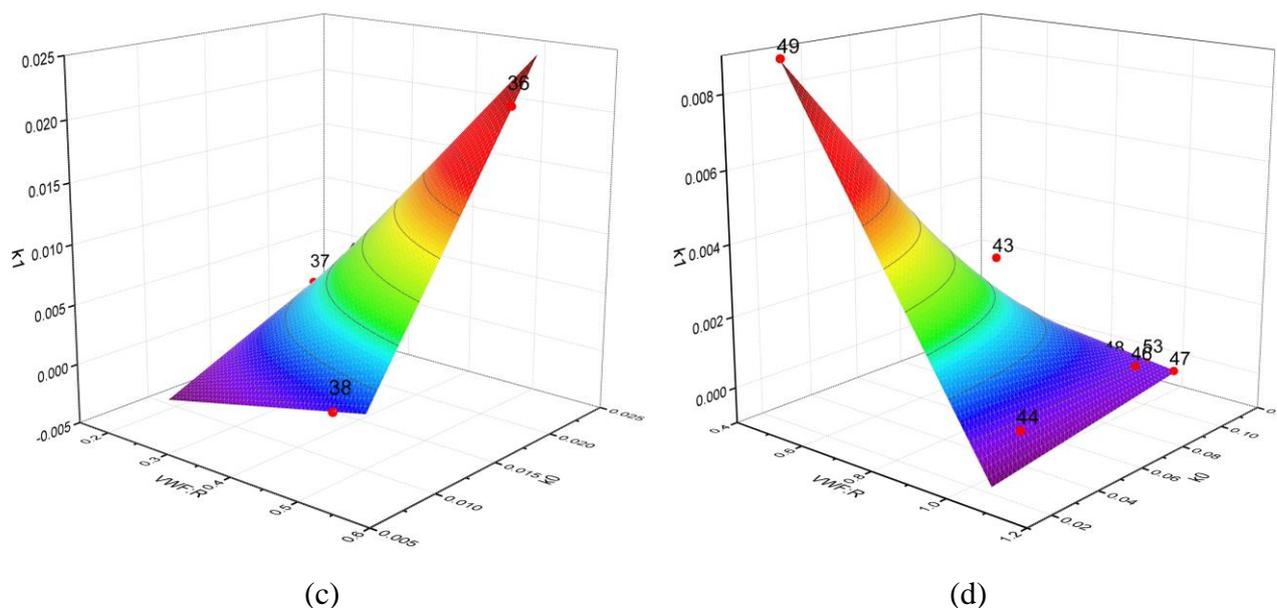


Figure B.11. Linear response surface with interactions considering HnonO (a), HO (b), 2B (c) and Vicenza (d) categories.

Table B.4. R^2 values for the different categories.

	R^2
HnonO	0.91985
HO	0.93521
2B	0.98516
Vicenza	0.97263

As table B.4 demonstrates, the value of R^2 is higher than 90% in all the considered categories, meaning that the fitting is really good. However, once the correlation for k_1 is substituted in MGM_1, the calculated value for the proteolytic parameter is completely incorrect in all the categories (table B.5).

Table B.5. Average k_1 estimated and calculated values in the considered categories.

	Average k_1 estimated	Average k_1 calculated
HnonO	0.001610986	0.00241648
HO	0.000383822	0.00065246
2B	0.003211634	-0.00256928
Vicenza	0.000552848	-0.00035932

This means that the parameters of the response surface do not allow a physical representation of the system.

The definition of the most suitable correlation for parameter k_1 has been reported in §4.1.2.

B.3 Investigation of the response surface for parameter k_0

The definition of the right correlation for the release parameter k_0 has been quite challenging. Indeed, as for the proteolytic parameter k_1 (§4.1.2 and §B.2), various basal state clinical quantities can be related to k_0 . Again, the collaboration with the medical school has been fundamental for determining the right response surface structure. The release of VWF from the endothelial cells is indirectly measured by the intraplatelet VWF (intra) and VWF:R quantities. In particular, intraplatelet VWF is the medical assay used for quantifying the amount of VWF contained in the endothelial cells. Therefore, it seems reasonable to investigate a correlation for k_0 as function of VWF:R and intraplatelet VWF. Results of the regression execution are reported in figure B.12 for HO, 2B and Vicenza categories. Intraplatelet VWF experimental data were not available for HnonO category.

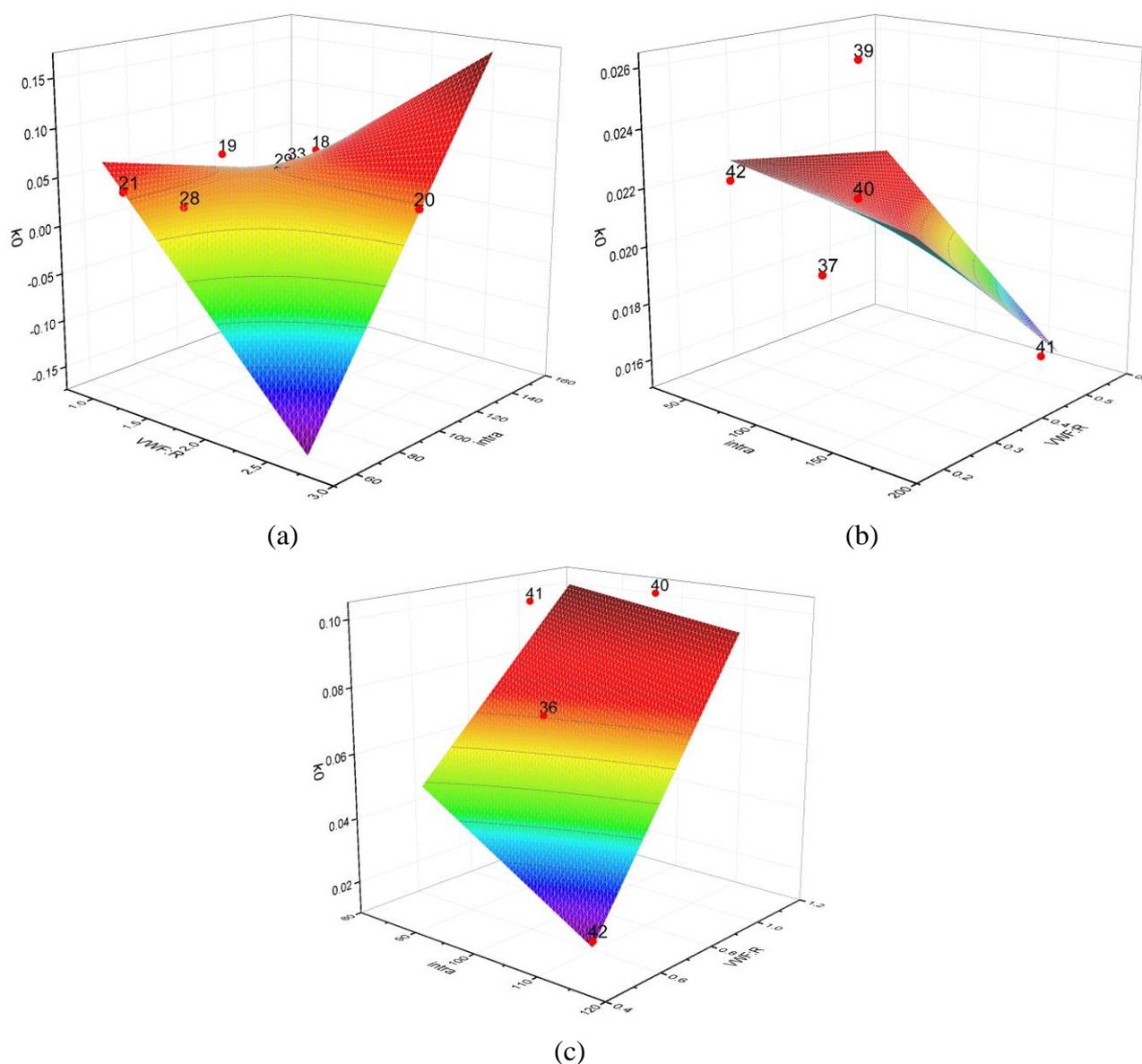


Figure B.12. Linear response surface with interactions considering HO (a), 2B (b) and Vicenza (c) categories.

As it is possible to visualize from the pictures reported in figure B.12 and from the values of the most important statistics in table B.6, the experimental data are not satisfactorily fitted by the response surface.

Table B.6. R^2 and \bar{R}^2 values for the different categories.

	R^2	\bar{R}^2
HO	0.63318	-0.0484
2B	0.66386	-0.39823
Vicenza	0.96374	0.8220

The correlation seems to work well only for Vicenza subjects, whereas for HO and 2B categories the fitting is really bad. This result is a clear demonstration that a more suitable correlation exists to represent the VWF release path.

The error made in the definition of the correlation for the release is conceptual. Indeed, differently from what happens for proteolysis and elimination of VWF, the clinical trials can measure only the amount of VWF (Q) released from the cells. Then, the rate of release can be obtained indirectly (§4.3). A correlation for Q has been tested as function of VWFpp ratio and VWF:R. Indeed, physiologically the measurement of the release is influenced by the elimination and by the amount of circulating multimers.

However, as figure B.13 and table B.7 illustrate, the result of the fitting is not completely satisfying for all the categories.

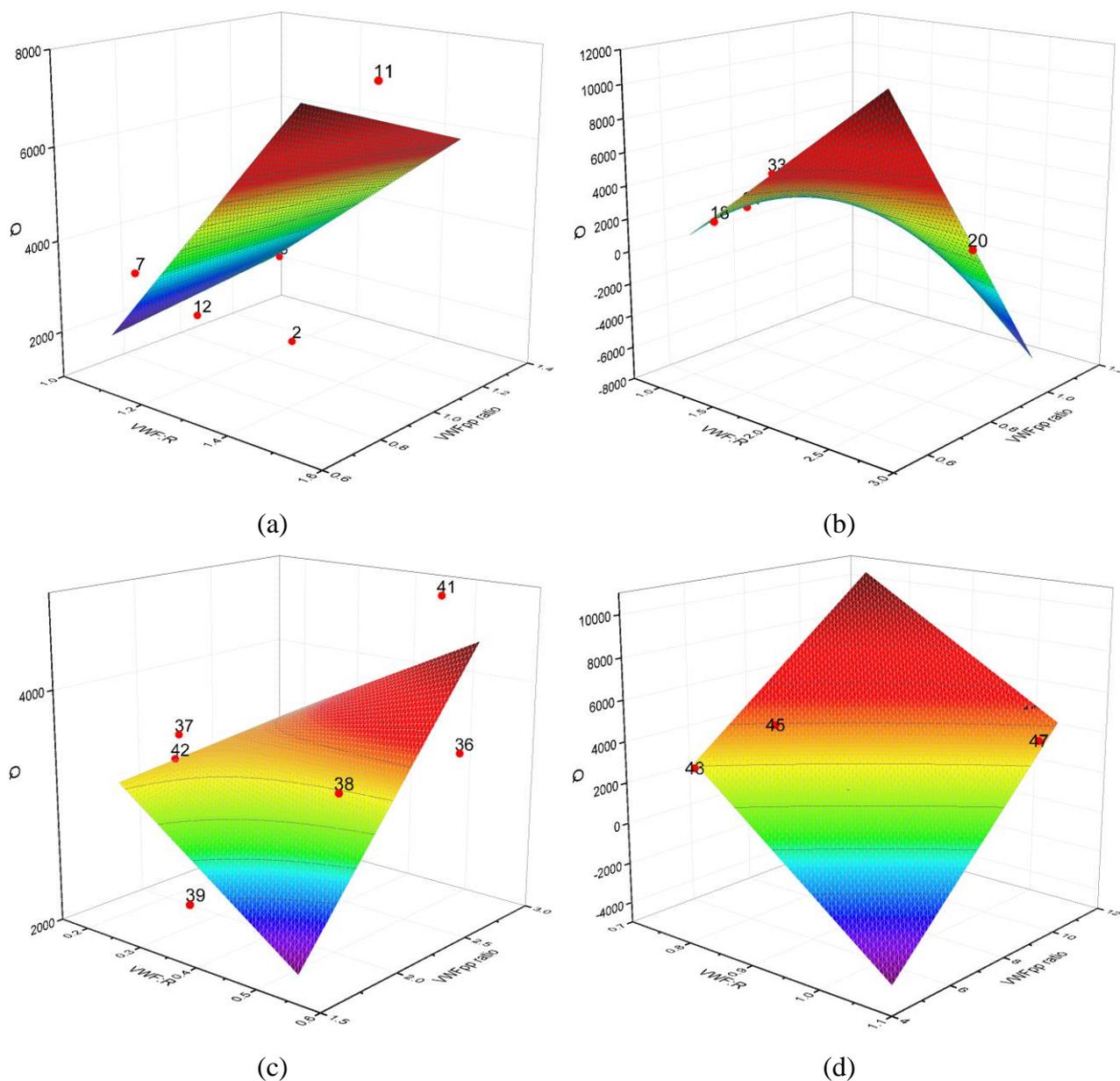


Figure B.13. Linear response surface with interactions considering *HnonO* (a), *HO* (b), *2B* (c) and *Vicenza* (d) categories.

Table B.7. R^2 and \bar{R}^2 values for the different categories.

	R^2	\bar{R}^2
HnonO	0.77254	0.29442
HO	0.52356	-0.27029
2B	0.73185	0.07121
Vicenza	0.88174	0.10984

Clearly, the statistics reported in table B.7 do not define a good fitting.

The correlations that have been found suitable for describing the VWF release path have been presented in §4.1.3.

Appendix C

C.1 In silico experimental data

To generate a random set of experimental data, at first the predicted values of the two model responses have been considered at each sampling time. Then, a random error has been added to the predicted values following equation c.1.

$$y_i = \hat{y}_i + \hat{e}_i \text{ with } i = 1, \dots, \text{ sampling point} \quad \text{c.1}$$

The random error has been calculated in Matlab[®], using the function *randn* and specifying zero mean and variance of four. The experimental data of the patients taken as a reference in each category are reported in table C.1. These values have been used to test the time reduced DDAVP protocol. The same procedure has been applied also to test the A-optimal and D-optimal protocols, whose generated experimental data are reported in figures C.2.a and C.2.b, respectively.

11	Time	prediction VWF:Ag	random error	experimental point	prediction VWF:CB	random error	experimental point
	0	104.5	1.0753	105.5753	155.2	1.0753	156.2753
	15	177.6479	3.6678	181.3157	228.13252	3.6678	231.80032
	30	225.1998	-4.5177	220.6821	275.1458	-4.5177	270.6281
	60	275.5614	1.7243	277.2857	323.8963	1.7243	325.6206
	120	302.2496	0.6375	302.8871	346.6452	0.6375	347.2827
	180	302.0264	-2.6154	299.411	342.48572	-2.6154	339.87032

32	Time	prediction VWF:Ag	random error	experimental point	prediction VWF:CB	random error	experimental point
	0	63.2	-0.8672	62.3328	93.8	-0.8672	92.9328
	15	117.34521	0.6852	118.03041	147.82715	0.6852	148.51235
	30	143.66428	7.1568	150.82108	173.8805	7.1568	181.0373
	60	161.33072	5.5389	166.86962	190.85645	5.5389	196.39535
	120	161.19516	-2.6998	158.49536	189.30708	-2.6998	186.60728
	180	155.20584	6.0698	161.27564	182.08286	6.0698	188.15266

37	Time	prediction VWF:Ag	random error	experimental point	prediction VWF:CB	random error	experimental point
	0	39.6	1.4508	41.0508	9.3	1.4508	10.7508
	15	72.25269	-0.1261	72.12659	41.031067	-0.1261	40.904967
	30	95.239914	1.4295	96.669414	61.735085	1.4295	63.164585
	60	121.67035	-0.4099	121.26045	81.59661	-0.4099	81.18671
	120	134.4222	-0.2483	134.1739	80.67713	-0.2483	80.42883
	180	127.80084	2.9794	130.78024	64.68337	2.9794	67.66277

45	Time	prediction VWF:Ag	random error	experimental point	prediction VWF:CB	random error	experimental point
	0	6.9	2.8181	9.7181	5.6	2.8181	8.4181
	15	82.85841	2.8344	85.69281	81.221756	2.8344	84.056156
	30	103.607574	1.343	104.950574	101.361916	1.343	102.704916
	60	95.83524	-2.415	93.42024	92.530495	-2.415	90.115495
	120	60.249924	1.4345	61.684424	56.1989	1.4345	57.6334
	180	37.558464	3.2605	40.818964	33.77802	3.2605	37.03852

Figure C.1. In silico experimental data to test the time reduced DDAVP protocol.

5					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	122.2000	131.7000	1.0753	123.2753	132.7753
18	162.6000	171.7165	3.6678	166.2678	175.3843
36	186.4600	194.6567	-4.5177	181.9423	190.139
60	203.8744	210.3768	1.7243	205.5987	212.1011
155	220.3806	218.8309	0.6375	221.0181	219.4684
172	220.9146	217.9103	-2.6154	218.2992	215.2949

20					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	54.50000	145.20000	1.4508	55.95080	146.6508
18	74.39568	164.98500	-0.1261	74.26958	164.8589
36	84.15627	174.5037	1.4295	85.58577	175.9332
60	88.86657	178.8371	-0.4099	88.45667	178.4272
155	81.99046	170.8617	-0.2483	81.74216	170.6134
172	80.09875	168.8741	2.9794	83.07815	171.8535

38					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	39.60000	9.30000	2.8181	42.41810	12.11810
18	77.53214	45.94188	2.8344	80.36654	48.77628
36	102.31800	67.65028	1.3430	103.66100	68.99328
60	121.67000	81.59582	-2.4150	119.25500	79.18082
155	131.56740	71.73276	1.4345	133.00190	73.16726
172	129.10210	66.94031	3.2605	132.36260	70.20081

48					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	6.10000	5.49000	1.7768	7.87680	7.26680
18	39.02970	38.39014	-2.2941	36.73560	36.09604
36	39.52139	38.84163	-2.1377	37.38369	36.70393
60	33.89742	33.11793	-1.6190	32.27842	31.49893
155	17.82841	17.08548	-5.8886	11.93981	11.19688
172	16.14233	15.40515	2.8768	19.01913	18.28195

Figure C.2.a. *In silico* experimental data to test the A-optimal DDAVP protocol.

5					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	122.20000	131.70000	-0.8672	121.33280	130.83280
20	165.91250	174.94700	0.6852	166.59770	175.63220
35	185.44340	193.69510	7.1568	192.60020	200.85190
55	201.17480	208.05700	5.5389	206.71370	213.59590
140	219.69590	219.43950	-2.6998	216.99610	216.73970
166	220.74860	218.31780	6.0698	226.81840	224.38760

20					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	54.50000	145.20000	0.6384	55.13840	145.83840
20	75.87579	166.44260	0.6257	76.50149	167.06830
35	83.79285	174.15550	-1.7298	82.06305	172.42570
55	88.39232	178.43770	-0.0601	88.33222	178.37760
140	83.69383	172.67300	-0.3298	83.36403	172.34320
166	80.75890	169.56500	1.2554	82.01430	170.82040

38					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	39.60000	9.30000	0.9778	40.57780	10.27780
20	80.84943	48.98615	2.0694	82.91883	51.05555
35	101.21150	66.74510	1.4538	102.66530	68.19890
55	118.63240	79.74600	-0.6069	118.02550	79.13910
140	133.25810	75.81951	0.5877	133.84580	76.40721
166	130.02460	70.33113	-1.5746	128.45000	68.75653

48					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	6.10000	5.49000	0.6504	6.75040	6.14040
20	39.74037	39.09612	-1.5099	38.19013	37.58622
35	39.70003	39.02229	2.7406	42.44063	41.76289
55	35.12236	34.41067	-3.423	31.69936	30.98767
140	19.54972	18.80308	-0.2045	19.34522	18.59858
166	16.70778	15.96841	-0.4829	16.22488	15.48551

Figure C.2.b. *In silico* experimental data to test the D-optimal DDAVP protocol.

To test the robustness of the redesigns DDAVP, experimental data with a variance of 36 have been generated and then used for parameters estimation. The *in silico* generated experimental data are reported in figure C.3.a and C.3.b.

5					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	122.20000	131.70000	3.2260	125.4260	134.9260
18	162.60060	171.71650	11.0033	173.6039	182.7198
36	186.46610	194.65670	-13.5531	172.913	181.1036
60	203.87440	210.37680	5.1730	209.0474	215.5498
155	220.38060	218.83090	1.9126	222.2932	220.7435
172	220.91460	217.91030	-7.8461	213.0685	210.0642

20					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	54.50000	145.20000	4.3524	58.85240	149.5524
18	74.39568	164.98500	-0.3783	74.01738	164.6067
36	84.15627	174.50370	4.2885	88.44477	178.7922
60	88.86657	178.83710	-1.2298	87.63677	177.6073
155	81.99046	170.86170	-0.7449	81.24556	170.1168
172	80.09875	168.87410	8.9382	89.03695	177.8123

38					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	39.60000	9.30000	2.9334	42.53340	12.23340
18	77.53214	45.94188	6.2082	83.74034	52.15008
36	102.31800	67.65028	4.3613	106.67930	72.01158
60	121.67000	81.59582	-1.8206	119.84940	79.77522
155	131.56740	71.73276	1.7632	133.33060	73.49596
172	129.10210	66.94031	-4.7237	124.37840	62.21661

48					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	6.10000	5.49000	1.9511	8.0511	7.4411
18	39.02970	38.39014	-4.5296	34.5001	33.8605
36	39.52139	38.84163	8.2218	47.7432	47.0634
60	33.89742	33.11793	-10.2691	23.6283	22.8488
155	17.82841	17.08548	-0.6135	17.2149	16.4720
172	16.14233	15.40515	-1.4487	14.6936	13.9565

Figure C.3.a. *In silico* experimental data with a variance of 36 to test the A-optimal DDAVP protocol.

5					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	122.20000	131.70000	-2.6016	119.5984	129.0984
20	165.91250	174.94700	2.0557	167.9682	177.0027
35	185.44340	193.69510	21.4704	206.9138	215.1655
55	201.17480	208.05700	16.6166	217.7914	224.6736
140	219.69590	219.43950	-8.0993	211.5966	211.3402
166	220.74860	218.31780	18.2095	238.9581	236.5273

20					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	54.50000	145.20000	8.4542	62.9542	153.6542
20	75.87579	166.44260	8.5032	84.3790	174.9458
35	83.79285	174.15550	4.029	87.8219	178.1845
55	88.39232	178.43770	-7.2449	81.1474	171.1928
140	83.69383	172.67300	4.3034	87.9972	176.9764
166	80.75890	169.56500	9.7814	90.5403	179.3464

38					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	39.60000	9.30000	5.3304	44.9304	14.6304
20	80.84943	48.98615	-6.8824	73.9670	42.1038
35	101.21150	66.74510	-6.4132	94.7983	60.3319
55	118.63240	79.74600	-4.857	113.7754	74.8890
140	133.25810	75.81951	-17.6657	115.5924	58.1538
166	130.02460	70.33113	8.6303	138.6549	78.9614

48					
Time	Prediction VWF:Ag	Prediction VWF:CB	Random error	Experimental point VWF:Ag	Experimental point VWF:CB
0	6.10000	5.49000	1.9152	8.0152	7.4052
20	39.74037	39.09612	1.8772	41.5772	40.9733
35	39.70003	39.02229	-5.1893	34.5107	33.8330
55	35.12236	34.41067	-0.1803	34.9421	34.2304
140	19.54972	18.80308	-0.9893	18.5604	17.8138
166	16.70778	15.96841	3.7662	20.4740	19.7346

Figure C.3.b. *In silico* experimental data with a variance of 36 to test the A-optimal DDAMP protocol.

C.2 Punctual sampling time optimization results

The results of the optimization executed for three experiment design A, D and E-optimal are reported in figure C.4. Subjects 11, 5, 15, 14 belong to HnonO category; subjects 32, 20, 30, 18 to HO category; subjects 37, 38, 41, 40 to 2B category, whereas patients 45, 48, 43 and 46 belong to Vicenza category.

11			5			15			14		
A-optimal [min]	D-optimal [min]	E-optimal [min]	A-optimal [min]	D-optimal [min]	E-optimal [min]	A-optimal [min]	D-optimal [min]	E-optimal [min]	A-optimal [min]	D-optimal [min]	E-optimal [min]
0	0	0	0	0	0	0	0	0	0	0	0
16.202	19.806	16.199	15.632	17.951	15.714	17.595	19.039	17.475	15	17.01	15
31.202	34.806	31.199	30.632	32.951	30.714	32.595	34.039	32.475	30	32.01	30
46.202	49.806	46.199	45.632	47.951	45.714	47.595	49.039	47.475	45	47.01	45
185	169.96	185	185	185	185	185	185	185	185	176.24	185
200	200	200	200	200	200	200	200	200	200	200	200

32			20			30			18		
A-optimal [min]	D-optimal [min]	E-optimal [min]	A-optimal [min]	D-optimal [min]	E-optimal [min]	A-optimal [min]	D-optimal [min]	E-optimal [min]	A-optimal [min]	D-optimal [min]	E-optimal [min]
0	0	0	0	0	0	0	0	0	0	0	0
15	15	15	15	15	15	27.798	32.114	27.796	15	18.355	15
30	30	30	30	30	30	42.798	47.114	42.796	30	33.355	30
152.56	45	152.72	45	45	45	57.798	62.114	57.796	45	48.355	45
167.56	138.02	167.72	173.21	160.19	173.34	185	185	185	159.73	139.6	159.74
182.56	153.02	182.72	188.21	175.19	188.34	200	200	200	174.73	154.62	174.74

37			38			41			40		
A-optimal [min]	D-optimal [min]	E-optimal [min]	A-optimal [min]	D-optimal [min]	E-optimal [min]	A-optimal [min]	D-optimal [min]	E-optimal [min]	A-optimal [min]	D-optimal [min]	E-optimal [min]
0	0	0	0	0	0	0	0	0	0	0	0
20.576	24.575	20.5	38.538	43.903	38.509	21.097	28.734	21.163	15	21.794	15
35.576	39.575	35.5	53.538	58.903	53.509	36.097	43.734	36.163	30	36.794	30
50.576	54.575	50.5	68.538	73.903	68.509	51.097	58.734	51.163	45	51.794	45
185	178.48	185	185	88.903	185	185	173.06	185	185	165.03	185
200	200	200	200	200	200	200	200	200	200	200	200

45			48			43			46		
A-optimal [min]	D-optimal [min]	E-optimal [min]	A-optimal [min]	D-optimal [min]	E-optimal [min]	A-optimal [min]	D-optimal [min]	E-optimal [min]	A-optimal [min]	D-optimal [min]	E-optimal [min]
0	0	0	0	0	0	0	0	0	0	0	0
15	15	15	15	15	15	15	15	15	15	15	15
30	30	30	53.018	53.866	53.038	30	30	30	58.721	30	58.758
75.364	74.445	75.146	68.018	68.866	68.038	66.807	64.152	66.848	73.721	45.491	73.758
90.364	89.445	90.146	83.018	83.866	83.038	81.807	79.152	81.848	88.721	60.491	88.758
105.36	104.45	105.15	98.018	98.866	98.038	96.807	94.152	96.848	103.72	75.491	103.76

Figure C.4. Punctual sampling time optimization for the selected subjects.